Assessing the risk of violent recidivism: Evaluating an alternative method of calculating the Violence Risk Appraisal Guide.

Suzanne J. O'Rourke

Doctorate in Clinical Psychology

University of Edinburgh

Declaration

I hereby declare that the work enclosed herein is my own except where otherwise stated.

Suzanne O'Rourke

Acknowledgements

Those to whom I owe thanks are many. First and foremost my supervisor, Dr. Paul Barret, for his inspiration, enthusiasm, constant support and his willingness to fight my battles for me (!). Dr. John Bogue for his academic supervision and advice. Dr. John Donnelly and Ms. Suzanne Allen for kindly sharing their data for the purposes of this study. My office mates, my parents and friends for supporting me through the many ups and downs of this last year.

Table of Contents

Table of Figures

Table of Tables

Abstract

The requirement for an accurate, time efficient and cost effective measure to predict violent recidivism in forensic hospital inpatients was highlighted by the Committee of Inquiry into the Personality Disorder Unit of Ashworth Special Hospital (Fallon et al., 1999) which identified a need for a 'consistent and standardised assessment protocol'. The Violence Risk Appraisal Guide (VRAG) is a such measure but in its traditional form requires the completion of the Psychopathy Checklist (PCL-R), a predominantly interview based assessment requiring a trained rater. The Child and Adolescent Taxon Scale (CATS), a records based assessment which is substantially faster to calculate without the requirement for either an interview or trained interviewer, has been proposed as a valid alternative to the PCL-R score within the VRAG (Quinsey et al., 1998). The present study in a UK Forensic inpatient setting, the State Hospital, examined the validity of calculating the VRAG using the CATS instead of the PCL-R in either its interview or records based form. It considers the influence of raters on reliability, the nature of differences between VRAG scores calculated using the CATS in comparison to the PCL-R, and the extent of measurement overlap between the CATS and PCL-R items. It was concluded that the calculation of the VRAG replacing the PCL-R score with the CATS is a valid and cost effective alternative but at the cost of some accuracy.

1 Introduction

1.1 The need to assess risk.

1.1.1 Introduction

There is a widely held belief that there is a link between mental illness and crime and that clear cut markers exist which indicate increased risk of dangerousness. The views of mental health professionals on risk of dangerousness are integral to detention under the mental health acts and increasingly sought (Campbell, 1995; Miller & Morris, 1986; Moore et al., 1983) during legal proceedings regarding initial disposal decisions and parole eligibility. Recent reviews of the care of mentally disordered offenders (Reed, 1992; Scottish Office, 1999), and persons with psychopathic disorder (Reed, 1994), the MacLean Consultation Paper (1999) and the enquiry into the Personality Disorder Unit at Ashworth Hospital (Fallon et al., 1999) have further highlighted the role of risk assessment. Rather than addressing community care and its associated risk management they address the post trial disposal of offenders, how they might best be managed within the judicial and hospital systems. Suggestions for alternative disposal options for mentally disordered offenders will serve to increase and further formalise the role of mental health professionals in assessing risk. Yet as MacLean points out "Any new disposal based on dangerousness ... would require an assessment of risk to be made, and may be difficult to justify unless the risk assessment procedure is sufficiently robust" (1999).

The present study reviews the literature on the ability of current clinical and actuarial measures to assess risk and evaluates a more cost effective method of calculating a promising actuarial measure, the Violence Risk Appraisal Guide.

1.1.2 The relationship between mental disorder and crime

The public perceive the mentally ill to be at a greater risk of committing violent acts (Borum et al., 1996; Bingley, 1997) perhaps reflecting the much publicised tragic failures of the recent shift towards care in the community (Bingley, 1997; North East and South East Thames Regional Health Authorities, 1994 - Regarding Clunis). A person who is dangerous has been defined as one who has "a propensity to cause serious physical illness or lasting psychological harm" (Home Office & DHSS, 1975) or who has "indicated by word or deed that he is more likely than most people to do serious harm, or act in a way that is likely to result in serious harm" (Walker, 1978). Risk in this context can be defined as the likelihood of a person acting in a dangerous way (Prins, 1996a), or of an event occurring (Snowden, 1997).

Is their perception accurate? Do the mentally ill represent a significant risk to the public? The conclusions of early reviews which suggested that there was no link between mental disorder and criminal behaviour (Bonita et al. 1998; Borum et al., 1996; Monahan & Steadman, 1983) have been superseded by the conclusion that the mentally ill do present a higher risk of re-offending than non-offenders but less than that of an offender population (Bonita et al., 1998; Home Office, 1993). When violent offending only is considered, prevalence rates are higher amongst the mentally ill and in turn there is a higher prevalence of mental disorder amongst those who commit violent crimes (Wessley, 1997; Borum et al., 1996; Monahan, 1992; Swanson et al., 1990; Swanson, 1994). However, this increased risk does not reflect a causal link between mental ill health and propensity to criminal behaviour but rather that a similar aetiology of offending behaviour exists in both the mentally ill and offender populations (Reed, 1997). Mentally disordered offenders and 'well' offenders tend to share demographic and historical characteristics (Monahan & Steadman, 1983) such as a history of prior criminality and social disadvantage (Chiswick & Cope, 1995). It has

therefore been suggested that the same predictors should be used to establish future potential risk in these populations and not symptomatology (Rice, 1997). The only exception to this may be the existence of current psychotic symptoms which may account for the increased lifetime risk of violence in schizophrenia (e.g. Lindquist & Allebeck, 1990; Wessely et al., 1994; Borum et al., 1996; Reed, 1997) and by definition those with psychopathic disorder (Reed, 1997).

1.1.3 The mental health acts and risk assessment

The relationship between mental disorder and risk of violence is legally recognised by current mental health legislation with the Mental Health Act 1984 allowing compulsory detention when it is considered 'necessary for his health and safety or the protection of others that he should receive such treatment' (MacLean, 1999). Such hospital orders may be accompanied by restriction orders where the patient is the responsibility of the secretary of state but their discharge is conditional allowing their recall for further treatment if necessary (MacLean, 1999). For discharge under either order the patient must no longer meet the criteria for detention, that is they must no longer present a risk to themselves or others, and decisions are guided by the principal that persons should be held under no greater security than is justified by the danger they represent (Reed, 1992). This has led to mental health professionals being responsible for a continual process of reassessment and monitoring of an individual's level of risk (Mulvey & Lidz, 1995; Reed, 1997). As they were recently reminded they must consider an individual's 'potential for dangerous behaviour' when discharging (NHS Executive, 1994). Care in the community brings with it rare, but considerable, risks for the public if release decisions are misguided, something which special enquiries into failures and media interest keep to the fore, perhaps biasing risk assessment through fear of liability (Diamond, 1974; Monahan, 1996). This is not an idle fear when

reconviction rates after discharge from special hospitals for serious offences are between 6 and 27% (Chiswick & Cope, 1995; Bailey & MacCulloch, 1994) rising to 34% over ten and a half years (Buchanan, 1998). In addition it has been recommended that mentally disordered offenders held under the mental health acts should have regular assessments of their needs which should be updated annually (Reed, 1992).

1.1.4 Proposed alternative sentencing options for mentally disordered offenders

The Mental Health (Scotland) Act 1984 attempted to incorporate persons with psychopathy, and without other diagnoses, within its remit by specifying that persons whose illness is manifest only by abnormally aggressive or seriously irresponsible conduct could be detained but only if treatment was likely to alleviate, or prevent a deterioration in, their condition (MacLean, 1999). As a result the House of Lords has upheld the argument that persons deemed untreatable could not be detained, and that preventative detention on the grounds of dangerousness was not allowable under the 1984 Act (MacLean, 1999). This ruling, in addition to creating much debate on the definitions of both personality disorder and treatability, has led to the acknowledgement that there is a 'hole' in our mental health legislation that may create difficulties for those sentencing offenders with personality disorder. This loophole in current legislation has recently been highlighted by a case that has received widespread press coverage concerning the release of a State Hospital patient on the grounds that his personality disorder is untreatable, an unpopular ruling that has provoked much public debate. To give mentally disordered offenders a determinate prison sentence allows them release at a fixed time without reference to their risk to society (Bacon, 1997; Duggan, 1997) whilst a hospital order may allow them release at any time they are considered no longer treatable and therefore detainable. A number of recent reports have

attempted to address this issue with proposals for alternative sentencing options for this population incorporating the assessment of risk.

The Reed Report (1994) highlighted the difficulties in assessing risk in persons with psychopathic disorder and recommended that, informed by research, 'a concise and practical set of guidelines should be prepared and widely distributed to those whose difficult task it is to assess risk and decide on proposals for discharge'. In addition they suggested the implementation of a 'hybrid order' where both a hospital order and prison sentence could be imposed simultaneously allowing the offender to be transferred to prison should they become untreatable when their discharge would pose a risk to public safety.

The issue of future risk is also central to the MacLean Consultation Paper (1999) whose remit is to examine the options for the management, treatment and sentencing disposal of violent and sexual offenders who present a continuing danger to the public. It asks the question of what weight should be given to the assessment of risk of continuing danger to the public when making sentencing and release decisions. In acknowledging the difficulties inherent in assessing dangerousness at time of sentencing it queries the concept of an 'indeterminate sentence' which would be based on risk. The duration of the sentence would be determined mainly by the perceived risk the offender posed to the public on release rather than by the offence they had committed and would be subject to a review process to determine an appropriate time of release (MacLean, 1999).

The most detailed proposal for alternative sentencing options for mentally disordered offenders, specifically those with personality disorders, was that detailed in the Report of the Enquiry into the Personality Disorder Unit at Ashworth Hospital (Fallon et al., 1999). It suggested that rather than a focus on psychopathy the real issue in offender disposal is

whether they represent a substantial risk of causing serious harm to others. Yet the dilemma here is that at sentencing neither a judge, nor anyone else, can predict at what stage in the future a mentally disordered offender will cease to represent a danger and therefore should be released. Fallon et al. (1999) suggest that offenders not in receipt of an indefinite or fixed sentence but deemed to be dangerous should be given a reviewable sentence which, after the end of a set tariff, would be renewable for up to two years at a time. Thus at the offender's earliest release date they could be assessed by a 'Reviewable Sentence Board' which would have within its power the ability to renew their sentence for a further two years or conditionally discharge the offender for two years after which they would be eligible for absolute discharge.

Initial assessment by the courts would continue as at present with preliminary psychiatric reports requested should it be deemed necessary to determine if the offender is suffering principally from a mental illness or mental impairment and should be dealt with under mental health legislation rather than through criminal legal process. However a third alternative would exist. Should the offender's principle diagnosis be considered a personality disorder the judge would instead request a 'Pre-Sentence Assessment' from an 'Assessment Team'. The Assessment team would be multi-disciplinary and psychiatry led and, using a 'consistent and standardised assessment protocol', seek to diagnose personality disorder and 'might also offer views on the dangerousness or otherwise of an individual'. In the light of this report the sentencing judge would consider whether the offender is both suffering from a personality disorder and 'presently poses a substantial risk of causing serious harm to others' should both questions be answered in the affirmative he would be obliged to pass a reviewable sentence.

In cases where either a reviewable sentence or a determinate sentence is passed on a person deemed to have a personality disorder, then a 'Clinical Management Assessment' would be

conducted. This would be a comprehensive multi-disciplinary assessment based in a prison special unit or in a high security mental hospital, using nationally agreed criteria, to identify the most appropriate disposal and care plan for the offender. When the offender's minimum tariff is reached the Reviewable Sentence Board would, in addition to having at its disposal all relevant records and evidence which the offender has chosen to present, know 'where possible the level of his current risk of causing serious harm to others'.

Should Fallon et al.'s (1999) proposals be introduced, the role of mental health professionals in assessing risk of violence in specific individuals would be substantially formalised creating a legal obligation for their input in:

- Pre-sentencing Assessments
- Clinical Management Assessments
- Evidence for the Reviewable Sentence Board

Reports during the current equivalents of these stages of offender management are already commonplace but these proposals for alternative sentencing options represent a substantial increase in the expectations of accurate risk assessment and in the weight that would given to their conclusions. In addition, it is specified that these assessments be based on UK wide, consistent, reliable criteria something which is lacking at present and that training in risk assessment should be compulsory for all staff in direct contact with the mentally ill (Royal College of Psychiatrists, 1996).

1.2 How do we assess risk?

Thus, the pressure on us to assess risk of violent recidivism is increasing. To enquire as to how we currently achieve this goal may appear straightforward but difficulty in defining the process quickly reveals that our current practice lacks clear structure, systems and consistency, even within professions. The most common current practice is clinical judgement, a method rooted in specialist knowledge that has recourse only to experience and is thus difficult to measure empirically. In an era when the terms 'evidence based medicine' and 'meta-analysis' are commonplace in health related fields, the ability of mental health professionals to give opinions based on such an assessment has caused concern (Meehl, 1997; Monahan, 1981).

The issues involve a balancing act between the loss of liberty for those we assess and the loss of safety for those in the community should our decisions be inaccurate. In acknowledgement of this fact, health professions have attempted to better define the process of clinical judgement, assess its efficacy and develop alternatives, or useful tools, in the form of actuarial measures.

1.2.1 A brief history of clinical risk assessment

The history of risk assessment is one characterised by tensions between those confident in the clinician's ability to judge for himself, a literature that casts doubt on this and the growing evidence that the answer may lie in actuarial measures. In the 1970s and 80s the long unquestioned concept of clinical judgement came under intense legal scrutiny when defendants argued that its unreliability was such that it did not constitute valid evidence at sentencing hearings (Monahan, 1996). Their arguments relied upon a series of studies in the

1970s which illustrated that at best clinical judgement was correct in only one third of cases and tended toward a lot of false positive errors (Monahan, 1981). Sweeping judgements stating ineffectiveness were made (Monahan, 1981; Marra et al., 1987) where clinicians and other professions questioned predictive abilities (Cohen et al., 1978) concluding error was inevitable (Megargee, 1976), that we could not and should not predict recidivism (Diamond, 1974) and indeed it was unethical to do so (Foot, 1990; Melton et al., 1987). Such was the persuasiveness of the conclusions from this first generation of research that even the American Psychiatric Association stated:

> "Neither psychiatrists nor anyone else have demonstrated an ability to predict future violence or dangerousness. Neither has any special psychiatric 'expertise' in this area been established."
> (American Psychiatric Association, 1974)

Similarly the American Psychological Association concluded:-

> "It does appear from reading the research that the validity of psychological predictions of dangerous behaviour, at least in the sentencing and release situation we are considering, is extremely poor, so poor that one could oppose their use on strictly empirical grounds that psychologists are not professionally competent to make such judgements."
> (American Psychological Association, 1978)

The results of these studies were accepted with little question and used to make damning judgements regarding detention on the grounds of mental health. For example the American Civil Liberties Union stated that 'it now seems beyond dispute that mental health professionals have *no* expertise in predicting future dangerous behaviour either to self or others. In fact, predictions of dangerous behaviour are wrong about 95 percent of the time' (Ennis & Emery, 1978a). In response to this overwhelming tide of criticism Monahan in an influential series of articles (Monahan, 1981; Monahan, 1984) argued that clinical judgement

was indeed sub-optimal (Monahan, 1984) and successful in only one third of cases (Monahan, 1981). However he based this assertion on long term community studies (Steadman & Cocozza, 1974; Cocozza & Steadman, 1976; Kozol et al., 1972; Steadman, 1977; Thornberry & Jacoby, 1979) and questioned the research to which he referred citing limited predictive methods, patient populations and time frames studied (Monahan, 1984). Otto (1994) also questioned the methodology of these '1st generation' studies noting inexplicit measures of risk, the limited release of patients, underestimates originating from records based outcomes and problematic definitions of violence. He suggested that perhaps a figure of one in two predictions being correct would be more accurate (Otto, 1992). Monahan (1984) too concluded that given different populations and time frames the ceiling on risk prediction may actually be closer to 50% than 5% if future research were to adopt different methodology (Monahan, 1984). His methodological recommendations included a shift to actuarial methods, both including clinical information and as aids to clinical judgement, that a broader range of risk factors be studied, including contextual, and that a wider population be examined including those requiring short term predictions in the community (Monahan, 1984). These recommendations have subsequently been referred to as forming the basis of a so called 'second generation' of studies in risk. Their results have been more promising but have failed to remove the high false positive rate and have not produced "clear cut means of achieving accuracy in prediction" (Webster et al., 1994; Borum et al. 1996). Although Monahan's recommendations have been largely welcomed the field of risk remains one in which great methodological difficulties exist.

1.3 Methodological difficulties in studying risk prediction

1.3.1 Defining violence

In any type of outcome study the definition of the 'outcome' itself is of pivotal importance. In the prediction of risk a difficulty arises because the definition of violence or dangerousness is far from unequivocal. While there may be little debate at the extreme end of the scale, that assault causing serious injury or death for instance should be included, questions about whether threatened violence, pushing, scratching or hair pulling should be included are more difficult. Studies have tended to adopt different definitions of violent or dangerous outcome making it difficult to compare their predictive value and altering their population base rates considerably. It should be noted that predictive accuracy is likely to increase the greater the scope of the outcome definition and that inaccuracy will result if clinicians fail to alter their predictions according to the severity of violence involved (Cunningham & Reidy, 1998). There is also a difference between likelihood and severity. A person may be deemed highly dangerous both if they are highly likely to engage in less severe violence and if they are at much lower risk of committing a violent act but one that, should they do so, will be of a very severe nature (Gottfredson & Gottfredson, 1988; Bjorkly, 1995).

Not only is defining violence a necessity to assess outcome, that is 'has a violent act been committed', but it is also an integral part of the prediction itself. If the clinician is predicting that an individual is, or is not, at risk for committing a violent act, what type of violent act are they referring to? It is a very different matter, in terms of release decisions, to consider somebody at risk of committing the violent act of a single angry thump in comparison to one who is deemed likely to take a life. Likewise context is important; certain situations may make the likelihood of re-offending greater and these should be specified. Violence is the

result of a complex inter-relationship between the individual, other persons and the environments in which they find themselves (Megargee, 1995). Prediction must therefore encompass both severity of violence and the context in which it is most likely to occur (Cunningham & Reidy, 1998). The severity of predicted violence also has implications for the clinician's predictive validity because it alters the population base rate (see 1.3.4).

1.3.2 The identification of recidivism and the opportunity to recidivate

Even if the issue of defining violence can be adequately solved the identification of such incidents is fraught with difficulty. Arrest records represent only a fraction of true offending, with many crimes remaining unreported (Mossman, 1994; Chiswick & Cope, 1995), unattributable, or with the offender remaining uncharged as they are considered better dealt with by the mental health system. Hospital re-admission records or inpatient records are of some use but are neither exhaustive nor undisputed.

Perhaps more important is whether, once a risk assessment has been made, the individual experiences the opportunity to recidivate. If a patient is considered to be at too great a risk of committing a violent act to be released from secure care, then the assessment of their risk may appear to be a false positive, they were predicted to re-offend but they did not (Litwack et al., 1993). It is the peculiar burden of this area of research that the persons in whom the assessment is of most interest are by definition those about whom clinicians will never receive feedback on their predictive performance. The ethical implications of releasing such persons to test predictions are of course too great to be contemplated. Indeed it can be argued that findings that victim death is a predictor of lower likelihood of future recidivism (Quinsey et al., 1998) could simply reflect persons who are not released to offend again. 'The most important and most certain 'predictions' of violence cannot be tested without

releasing from confinement precisely those individuals who most clearly pose a serious risk of committing serious violence if released.' (Litwack et al., 1993).

Yet two 'naturalistic' experiments have taken place where, due to legal decisions, patients previously considered 'dangerous' have been released without psychiatric approval providing complete release samples: the Baxstrom and Dixon populations (Steadman & Coccozza, 1974; Thornberry & Jacoby, 1979). Unfortunately both studies, as a result of their naturalistic nature, suffered the same methodological flaws which limit their usefulness. The samples were atypical of psychiatric populations being much older, but more importantly no explicit opinion on their dangerousness was established prior to their release, they were assumed all to be considered dangerous by virtue of their incarcerated status. Thus their most useful contribution was the finding in both samples of a lower rate of recidivism than was expected 20% and 24% (Steadman & Coccozza, 1974; Thornberry & Jacoby, 1979) further clarifying the difficulties of avoiding false positive errors in low base rate behaviours (Webster et al., 1996).

Are clinicians overcautious regarding release decisions, do they maintain patients at a higher level of security than necessary? In making release decisions clinicians must weigh the relative moral costs of detaining a patient in a secure setting unnecessarily (Pollock & Webster, 1990) in comparison to the cost of a violent act in the community. In addition to moral and human costs, the financial costs of a patient remaining in a higher security facility than necessary, and the health and legal costs of a violent act, are substantial. Thus the clinician's dilemma involves a precarious balancing act between doing the best for their patients whose liberty is at stake and the well-being of the general public (Chiswick & Cope, 1995; Diamond, 1974). It is also questionable whether, even if one's prediction is accurate, it is morally acceptable to detain someone for an act they have yet to commit (Palermo et al., 1991; Bingley, 1997; Miller & Morris, 1988).

1.3.3 Reporting outcomes

Traditionally studies reflect this dilemma in the reporting of the sensitivity and specificity of clinicians' judgements with particular favourites being their positive and negative predictive value and the immensely dilute measure of percentage of correct judgements. The aim is of course to find a measure that possesses both a high sensitivity and high specificity, but in reality high accuracy in one tends to be at the cost of low accuracy in the other. Indeed it has been argued that such a quantitative approach is inappropriate in the field of risk assessment because clinicians are referring to probabilities rather than dichotomous outcomes. When suggesting an individual is at high risk of re-offending and therefore should not be released a clinician is not stating that they *will* re-offend, rather that they are at an unacceptably high risk of re-offending (Litwack et al., 1993). This issue is further complicated by the fact that clinicians are dealing with subjective probabilities. It is not possible to compare the concept of 'medium risk' across clinicians because 'medium risk' does not have numerically defined criteria.

1.3.4 Base rates

The population base rate refers to the prevalence of a given behaviour in a population over a set time period, so in the present discussion it would refer to the number of violent acts in a given population during a given period. It will therefore change according to the definition of violence adopted and the population sampled. An understanding of this concept and its implications for risk prediction is of pivotal importance. It has been considered the most important information necessary for accurate prediction (Monahan, 1981) and its neglect has been referred to as one of the most significant errors made by mental health professionals (Monahan, 1981; Smith, 1993).

The population base rate for a given act represents a confounding factor for the traditional indices of measures of success discussed previously. For example the percentage of accurate predictions commonly quoted fails to take into account both base rates and clinical bias toward false positive errors (Mossman, 1994). The lower the base rate the more difficult it is to predict a behaviour accurately without inflating false positive errors excessively. The closer the base rate is to 50% the more accurate judgement is likely to be (Meehl & Rosen, 1954). High base rates lead to higher positive predictive value whilst lower base rates lead to higher negative predictive values (Bjorkly, 1995). For a very low base rate the best prediction is to predict no recidivism whilst with high base rates the opposite is true (Limandri & Sheridan, 1995). The margin to which the true positive rate exceeds the base rate can be a good indicator of the predictor's incremental validity (Otto, 1992). Of course, as base rates for a given act depend on the definition of the act itself, they can be artificially inflated by broadening the definition used, for example including violence of a milder severity. This has been suggested as one way in which amongst the mentally ill the false positive rate can remain acceptable and correct classification rates exceed those of first generation research (Otto, 1992). Yet the practical utility of predicting a broad range of violence is questionable as one would not wish to deny liberty to someone on the basis they of recidivism that might only constitute a verbal threat for instance. However for the same reason that we are unable to receive feedback on predictions in higher risk samples the fact that these persons are not released also denies us knowledge of a base rate of violent recidivism for this population. This is a perennial dilemma for risk validation studies.

1.4 Clinical Judgement

Clinical judgement refers to the examining clinician's considered opinion in the light of his examination and appraisal of other sources using his professional expertise. As previously discussed, early research on risk prediction portrayed clinical judgement in a very poor light, most notably with the assertion that it was correct in only one third of cases (Monahan, 1981; Webster et al., 1994). Yet this has not prevented a pro-clinical judgement lobby arguing that risk assessment should still be rooted in clinical skills (Greenland, 1980; Prins, 1986b, 1988), that no other measurement is available (Reed, 1997), that negative interpretations of the literature are seriously misguided (Litwack et al., 1993) and that focusing merely on predictive expertise neglects the broader contribution health professionals have to make (Pollock & Webster, 1990). Perhaps more evidence based arguments for the efficacy of clinical judgement may be those that suggest clinical judgement is more accurate in certain circumstances. These might include instances when short term predictions are required (Meehl, 1973; Otto, 1992), when disinhibitory or situational factors are implicated (Pollock, 1990), when persons possess a history of violence (Otto, 1992), in contexts similar to the outcome context (Otto, 1992), or when based on empirical evidence and developed according to rigorous standards (Governor's Task Force, 1994). Such arguments have led commentators to conclude that when certain guidelines are followed clinicians can accurately predict dangerousness in some situations (Cohen, 1996; Litwack et al., 1993) and that at least half of their *short term* predictions are accurate (Otto, 1992). The North American legal system seems to be in some agreement allowing clinical judgement as legitimate evidence and ruling against defendants who have argued for its dismissal as inaccurate (Webster et al., 1994; Litwack et al., 1993).

Critics of clinical judgement focus primarily on the existence of bias and inconsistency in clinical judgement. Critics argue that rather than consistently applying a context free

algorithm (Mulvey & Lidz, 1995) clinicians fail to allow for systematic biases employing instead a cognitive set of shortcuts and representations making insightful, rather than statistically sound, judgements' (Turk & Salovey, 1988). One of the most consistent biases that clinicians are prey to is a tendency to false positive errors, that is they consistently over predict violence, although perhaps this is understandable (see 1.1.3) (Cunningham & Reidy, 1998; Otto, 1992). Bias in clinical judgement stems from a wealth of sources (Shah, 1978). Ziskin's (1995) extensive discussion of this matter cites the under-utilisation of base rates, hindsight bias, confirmatory bias, anchoring effects and over reliance on salient data. Monahan (1981) points to a lack of specificity in defining criterion, failure to incorporate environmental information and again the ignoring of statistical base rates. Whilst clinicians have access to the historical and demographic data that actuarial measures utilise they lack the ability to distinguish the valid from invalid, to weight them optimally (Dawes et al., 1989) or grasp complex information (Ziskin, 1995). Contributing to this are the illusory correlations that occur when feedback on predictions points to variables being valid, or predictive abilities to be accurate, yet by nature such correlations must lack reliability due to the very limited and skewed nature of samples when feedback is received (Webster et al., 1994; Cunningham & Reidy, 1998; Shah, 1978; Ziskin, 1995; Monahan, 1981). The feedback we receive is at best inadequate to aid our learning and at worst gives us a false sense of confidence and creates faith in factors which may not warrant it. Clinical experience is not necessarily an advantage (Cunningham & Reidy, 1998), including in comparison to the educated layman (Quinsey, 1979; Quinsey & Ambtman, 1979) and may even be a disadvantage (Dawes et al. 1989).

Pro-actuarial commentators have not been kind to clinical judgement calling mental health professionals 'poor judges' (Harris, 1997), with 'very limited ability' (Mulvey & Lidz, 1995) and 'no expertise' that are wrong 95% of the time (Ennis & Emery, 1978b). Suggesting they were 'doomed to failure' (Otto, 1992), that it was 'established fact that ...even under the best

conditions...prediction(s)...are wrong in at least 2 out of every three cases' (Barefoot v Estelle, 1983; Otto, 1992), that 'expert testimony is not expert at all' (Faust & Ziskin, 1988; Quinsey and Ambtman, 1979), that there is a 'long standing, wide consensus...that health professionals are unimpressive in their ability to predict dangerousness' (Mulvey & Lidz, 1995), and suggesting we 'abandon the attempt' (Stone, 1985). Claims to special expertise have been unsupported in study conditions (e.g. psychiatrists vs. teachers (Quinsey & Ambtman, 1979) and our faith in our abilities does not constitute evidence (Meehl, 1997; Kleinmuntz, 1991; Ziskin, 1995), although confidence in specific judgements may be related to accuracy over short periods (McNeil et al., 1998). Yet without recourse to specialist expertise the very pertinent problems of detecting malingering, identifying inaccuracies in file data and the adjustment of risk prediction to encapsulate dynamic variables seems impossible (Pollock & Webster, 1990).

1.5 Actuarial Judgement

Thus mental health professionals are left in a quandary. Their clinical judgement lacks predictive validity and yet the legal system and mental health legislation persists in not only requesting their opinion but actually building that opinion into the system itself (Miller & Morris, 1988). During the last decade research efforts have focused on actuarial methods in an effort to either replace, or provide a useful adjunct to, traditional clinical judgement. Actuarial prediction consists of quantitative models derived from the isolation of optimally predictive variables (Mulvey & Lidz, 1995). The concept relies on the idea of group membership and that individual's share the qualities of these groups, thus if a certain group is considered more dangerous the individual will in turn be considered such in proportion to his similarity to the said group (Grisso & Appelbaum, 1992). Group statistics have been deemed not only relevant and applicable to individuals but 'indispensable' (Cunningham &

Reidy, 1998) and are considered to lead to better release decisions in insanity acquittees with the proviso that the persons on whom the model are applied are similar to the development sample (Rice, 1997).

Actuarial prediction of violent recidivism has been considered superior to clinical judgement by many (Goldberg, 1970; Bonita et al., 1998; Rice, 1997; Miller and Morris, 1988; Mossman, 1994; Noak, 1997) and remains so when cross validated on new, rather than developmental, samples (Copas & Whiteley, 1976). This is consistent with the experimental literature which has consistently found in favour of actuarial judgement in many fields (Grove & Meehl, 1996; Goldberg, 1970). Rice (1997) summarised "...in almost every situation in which they have been studied, actuarial predictions have outperformed unaided human judgement (Meehl, 1954, 1986, 1996). Meehl (1986) concluded, 'There is no controversy in social science that shows such a large body of qualitatively diverse studies coming out so uniformly in the same direction as this one' (Rice, 1997; Polythress, 1992) and 'the mechanical method is almost invariably equal to or superior to the clinical method' (Grove & Meehl, 1996). It is significantly cheaper than the clinical method and ideally suited as a screening tool in addition to being an explicit method allowing both availability to others and empirical research (Dawes et al., 1989).

Yet actuarial judgement has not been without its critics (Greenland, 1980). Doubts have been voiced regarding its value in individual cases (Snowden, 1997; Litwack et al., 1993; Chiswick & Cope, 1985; Snowden, 1997) (although countered as ignoring the basic laws of probability; Dawes et al., 1989), its reliability without complete base rates (Litwack et al., 1993) and its inability to consider the possible consequences of a decision (Litwack et al., 1993). Critics have suggested that it is reliable only for extremely high and low risk groups (Pollock & Webster, 1990), that violence is simply too complex to be predicted by a single test (Megargee, 1995) and that criterion variables are too poorly defined to allow reliability

(Chiswick & Cope, 1985). Even its proponents acknowledge that it is not infallible and that measures require periodic re-evaluation and careful consideration before their application to new settings (Dawes et al, 1989). For some, actuarial judgement is considered useful only in circumstances when clinical judgement is more obviously struggling, for example when there is little consistency or pattern to offending (Pollock, 1990) or merely as a useful adjunct to the clinical decision making process (Litwack et al., 1993) especially to avoid false negatives (Cunningham & Reidy, 1998; Goldberg, 1970).

1.6 Literature review

The inclusion criteria for the following review were that studies examine the accuracy of clinical judgement in offenders who subsequently had the opportunity to recidivate violently in a community setting. Therefore certain categories of studies have been omitted including those predicting inpatient violence (Yesavage et al., 1982; Werner et al., 1983; Rofman et al., 1980; McNeil et al., 1988; McNeil & Binder, 1987, 1991, 1994; Lansing et al., 1997; Krakowski, 1994; Apperson et al., 1993; Bjorkly, 1994; Blomhoff et al., 1990; Convit et al., 1988; Janofsky et al., 1988; Kirk, 1989) and within prison violence (Cooper & Werner, 1990; Jones et al., 1981; Selby, 1984; Swett & Hartz, 1984). Also omitted are studies of juveniles (Forth et al., 1990), those not specifically of violence (Hart et al., 1988; Hoffman & Beck, 1974, 1985; Russo, 1994) and those examining physiological predictors of recidivism (Howard & Lumsden, 1997; Kandel et al., 1989). These inclusion criteria were to allow the review to draw comparisons with the stated aims of the VRAG to predict violent recidivism after release into the community. The following literature review is the result of searches of the Medline and Psychlit databases from 1966 to 1999. All review and primary articles were consulted and the papers to which they referred checked to see if they met inclusion criteria.

It quickly becomes apparent from the preceding history of debate regarding the efficacy of clinical judgement that to give greater weight to methodologically flawed studies than they deserve can serve to create and perpetuate misinformation. It is with this in mind that I approach the following literature review following the guidance of Sackett et al. (1985) regarding the methodological criteria for outcome studies.

Sackett et al. (1985) suggest that one should consider six criteria when critically evaluating outcome trials to properly allow one to weight the strength of their evidence.

(1) Was an inception cohort assembled?

Were the sample identified at an early and uniform point in the course of their 'disease'? It is important that the population be at a similar stage of their 'disease career', in our case offending history, because the prognoses, or likely recidivism, of a first offender in comparison to a repeat offender at the peak of his career or one who has 'burnt out' may be very different. Ideally a sample should comprise only those at a similar and very early stage so that it is not subject to bias introduced by omitting one time only offenders or those whose offences were of a magnitude that they were not released to enter the system on a subsequent occasion. Such an early and uniform point is difficult to identify in the risk literature but could be at time of conviction for first violent offence or at time of first hospitalisation on the grounds of dangerousness. Of course both of these would still be subject to some bias and a population cohort study from pre-adolescence would be the method of choice.

Was information gathered retrospectively?

By definition, retrospective studies are skewed toward repeat offenders as this population has a greater chance of inclusion in such a study than one off offenders. The retrospective isolating of variables that predict membership of this group is also problematic since to discover, for example, that the majority were arrested before the age of sixteen years may appear a predictive variable but fails to denote the proportion of all persons arrested under the age of 16 years who subsequently commit a violent crime.

Were subjects with some outcomes excluded?

As previously discussed with regard to base rates (see 1.3.4) the literature on risk is almost without exception limited to samples in which those persons predicted to be at most risk of violent recidivism are omitted, because of this prediction, and the prediction itself therefore never tested. This may work both in favour and against the arguments for our ability to predict risk. Perhaps those never released are those amongst whom we are most accurate at predicting risk and their exclusion therefore reduces our 'average' or perhaps the reverse is true and they represent a high rate of false positive predictions.

(2) Was the referral pattern, for entry into the sample, described?

Access to institutions is governed by differing pathways which possess many gateways each of which determine who remains en route for entry to a particular system and who is side-tracked into another. Of central importance in any outcome study is whether the population to whom you would like to apply the results is similar to that in which the study was conducted. It is therefore of great importance to understand how subjects reached the institution, the characteristics they may therefore possess, and how that may have biased their outcome.

Centripetal bias: Difficult cases are more likely to be referred to specialist centres. Therefore comparison between results from a young offenders institution and a maximum security national institution are likely to produce different base rate recidivism and different levels of predictive accuracy.

Popularity Bias: Professionals may be more likely to keep track of, or monitor, more interesting cases. In recidivism studies more interesting or difficult cases may be at a reduced risk of being lost to follow up once in the community both because of increased professional interest amongst after-care services but also should they be considered at the 'riskier end' of those released.

Re-referral filter bias: The nature of referral to an institution may create bias in the population sample it receives. Is the facility seen as the last resort when others have been exhausted? Or is it perceived as taking 'a certain type' of offender?

Diagnostic Access bias: Do factors such as geography or socio-economic class affect referral to the institution? If the courts or other medical facilities perceive a facility as 'difficult for the non street wise' would they reconsider the disposal of someone who was particularly young, naive or middle class? If referral is partially determined by geographical catchment area the population of this area needs to be described to facilitate comparison with other samples.

(3) Were all members of the inception cohort traced?

Of those offenders counted as entering the study, for example those on whom predictions were made and who were then released into the community, were all followed up successfully? The reason someone is lost to follow up often has a bearing on their outcome such that failure to account for all persons originally entered into the

trial may bias its results. If for instance a number of persons left the follow up area, moved country perhaps, and were lost to follow up this may indicate a wish for a new, crime free, start, alternatively it may indicate a wish to avoid parole conditions and benefit from the freedom to offend the anonymity a new location can provide.

(4) Were objective outcome criteria developed, used and consistently applied?

To effectively study the prediction of violent recidivism its definition must be developed and applied consistently. This is not as clear cut as it appears as even if arrest records are the follow up source the decision must be taken of whether charges count or only actual convictions, and exactly which crimes will be considered 'violent'. The picture is further muddied by the fact that not only are arrest records vast under-estimates of criminal activity but that in the case of the mentally disordered, often they are re-admitted to psychiatric facilities rather than facing new charges.

(5) Was the outcome assessment blind?

Whilst the concept of a 'double blind trial' is the most basic tenet of every drug trial its importance in other studies should not be forgotten. Was the researcher combing records for signs of recidivism 'blind' to the prediction made prior to the offenders release? Whilst ostensibly record trawling is an objective measure researcher expectation can lead to more or less detailed searching according to their bias. If one is not expecting to find signs of recidivism one may not look as hard for it as one might if one were surprised not to have found it.

(6) Did adjustment take place for extraneous factors?

Despite a researcher's best efforts certain biases or extraneous factors may exist within their sample that make it unique or lacking in generalisability. These additional factors can be controlled for statistically but an additional, more robust, method to examine the validity of one's hypothesis is to examine it in a new sample distinct from that on which it was developed. This is especially applicable to studies of actuarial prediction of recidivism where it is necessary for any actuarial instrument to be validated on a new sample other than that on which the measure was developed (Dawes et al., 1989).

1.6.1 Studies of clinical judgement

1.6.1.1 Predicting recidivism within 12 months using clinical judgement

The literature review revealed two articles which examined the efficacy of clinical judgement when predicting violence during a period of up to one year (Table 1).

Neither study had identified an inception cohort at an early and uniform point, one taking emergency room referrals (Lidz et al., 1993) and one emergency community referrals (Levinson & Ramsay, 1979). Both suffered from substantial referral biases and offered little detail regarding the route of subjects into the study facility. The sample of Levinson and Ramsay (1979) could not adequately describe the likely influences upon the referral of its subjects because it was based on home visits to those reported by members of the public to be causing concern. Of those referred the decision regarding whom should receive a home visit and thus become eligible for participation appears to have been clinical. Although the formal 'gateways' of referral can be a source of bias their complete absence is likely to have

Table 1: Studies of clinical risk prediction with a follow up of up to one year

Authors	number Mean age % Male	Population	Time to follow up	Predictive Method	Outcome Definition	Base Rate	AUC
Levinson & Ramsay, 1979	53 N/K N/K	Referrals from community for crisis intervention	3 months	Clinical Judgement	Dangerous behaviour reported in follow up interview	25%[a]	0.57[a]
Lidz et al. 1993	714 28 yrs 60%	Hospital emergency room pts Predicted aggressive group: 15%-Schizophrenia 16%-Affective disorders 41%-substance abuse Comparison group: 16%-Schizophrenia 30%-Affective disorders 7%-substance abuse	6 months	Clinical Interview using 10 point scale reflecting potential patient aggression	Physical assaults or threats with a weapon noted from self report, information from others or records	45%[a] 36%[a]	0.66[a]

Table columns reflect the statistical information available.
[a] Figures not quoted in paper. Source is Mossman (1994).
[b] Figures not quoted in paper. Source is Bjorkly (1995).
AUC= Area under the ROC curve.
Pts= patients
Inpts = Inpatients
Sjs = subjects

contributed even greater bias as referral was dependent on the knowledge and concern of neighbours. Lidz et al.'s (1993) study approached consecutive referrals to an emergency room to request their participation and offered them a cash fee. Seventy-nine percent agreed to participate and the authors acknowledge that the sample was skewed toward lower socio-economic groups because of the financial incentive. Neither study achieved complete follow-up of all subjects. Levinson and Ramsay (1979) managed to follow up only 53% of their subjects, the majority of the remainder being described as 'could not be contacted' but assert that missing subjects were not considered more dangerous at initial assessment. Lidz et al. (1993) fared little better with 23% of subjects either dropping out or lost prior to the completion of follow up and a further 13% omitted as it was not possible to find matching controls.

Lidz et al. (1993) defined violent outcome as the laying of hands upon another with violent intent or threatening with a weapon according to either official records, self or collateral report. Levinson and Ramsay (1979) describe using two categories of violent outcome to which patients were classified according to open and closed questions asked of the initial referrer (member of public) The questions are not specified. In classifying outcome neither study mentions whether researchers were blind to the patient's predicted outcome.

The two studies come to opposing conclusions. Lidz et al. (1993) conclude that clinicians could predict violence at a better than chance level during the next six months with an area under the curve (AUC) of 0.66 and false positive rate of 47%. The AUC refers to the area under a Receiver Operating Characterisitics (ROC) curve, a plot of sensitivity against 1-specificity, an indication of predictive accuracy ranging from 0 to 1.0, where 1.0 indicates perfect accuracy and 0.5 indicates chance levels. Levinson and Ramsay (1979) found no significant relationship between clinician prediction of violence and outcome at three months. Yet this was methodologically the poorer study, with a much smaller sample size

and, perhaps most importantly, the clinicians in question were graduate mental health associates with no formal clinical training.

1.6.1.2 Predicting recidivism in one to three years using clinical judgement

The literature search identified five studies examining the efficacy of clinical judgement to predict violent recidivism over a period of two or three years (Table 2). None of the four studies identified members of their sample at an early and uniform stage in their careers with three (Sepejak et al., 1983; Webster et al., 1984; Cocozza & Steadman, 1976) taking patients at pre-trial assessment, one examining offenders prior to release from long term psychiatric care (Werner & Meloy, 1992) and one following up released patients (Steadman & Cocozza, 1978). Two studies were affected by the exclusion of certain patients, that of Werner and Meloy (1992) whose subjects were selected by virtue of their being considered for release and Steadman and Cocozza (1978) where only 60% were released into the community. Two studies overcame this difficulty by following patients whether their disposal was to another institution or the community (Sepejak et al., 1983; Webster et al., 1984; affecting the relevance of their results to the present discussion) and in one this matter was unclear (Cocozza & Steadman, 1976). Two studies were based in the same institution, a pre-trial brief assessment unit (Sepejak et al., 1983; Webster et al., 1984), two received their referrals from courts (Cocozza & Steadman, 1976; Steadman & Cocozza, 1978) and the remainder gives no detail except that patients were resident in a long term psychiatric facility (Werner & Meloy, 1992).

The studies of Werner and Meloy (1992), Cocozza and Steadman (1976) and Steadman and Cocozza (1978) achieved complete follow up but those of Sepejak et al. (1983) and Webster et al. (1984) have adopted the same rather unusual stance. Their analyses include only those

Table 2: Studies of Clinical Risk prediction with a follow up of two to three years

Authors	number Mean age % Male	Population	Time to follow up	Predictive Method	Outcome Definition	Base Rate	Chi²	AUC	True +ve	False +ve	True -ve	False -ve	Correct
Cocozza & Steadman, 1976	257 N/K 100%	Psychiatric pts	3 years	Assumption of dangerousness due to inpt status	Assaultiveness	15%[a]		0.48[a]					
Sepejak et al., 1983	364 N/K N/K	Forensic brief assessment unit pts	2 years	Four point clinical judgement scale	Criminal & institutional behaviour rated on 11 point scale	46%[a]		0.64[a]					
Steadman & Cocozza, 1978	154 N/K N/K	Indicted felony defendants found incompetent to stand trial	Unclear	Clinical judgement	Assaultiveness	6%[a]	NS						
Webster et al., 1984	158 Majority 16-30 yrs N/K	Forensic brief assessment unit pts for whom recidivism data found	2 years	Assessment interview and Dangerous Behaviour Rating Scheme	11 point scale of degree of dangerousness from records	24%[a]		0.68[a]	76%[b]	64%[b]	24%[b]	35%[b]	43%[b]
	217 N/K 88%	Forensic brief assessment unit pts, entire sample				36%[a]							
Werner & Meloy, 1992	50 38 yrs 92%	State hospital pts	2 years	The forensic adaptation of the BPRS clinical interview	Violent acts warranting clinical or law enforcement	0%					100%	0%	100%

[a] Figures not quoted in paper. Source is Mossman (1994); [b] Figures not quoted in paper. Source is Bjorkly (1995); AUC= Area under the ROC curve. Pts= patients; Inpts = Inpatients; Sjs = subjects; NS = Not signifcant; Yrs = years.

patients on whom they could find at least one piece of follow up data during their search of criminal and psychiatric records. That is, if no evidence was found that patients had recidivated (violently or non-violently) they were excluded from the outcome analysis 'based on the uncertainty as to whether these subjects remained incident free' (Sepejak et al., 1983). This limits the ability to identify a true negative rate to those patients who were guilty of non-violent recidivism and alters the base rate in an upward direction. Yet it was these latter two studies which had best defined their outcome criteria assigning scores on a scale of 1 to 11 in comparison to merely describing 'assaultativeness' (Cocozza and Steadman, 1976; Steadman & Cocozza, 1978) or 'a violent act' (Werner & Meloy, 1992). The studies of Sepejack et al. (1983) and Webster et al. (1984) were also the only two which addressed outcome researcher 'blinding' to predictive status by using external, independent researchers to rate follow up data. However it is not clear if this applied only to their rating of subjects on the outcome scale or if it also applied to the collection of follow up data, both being preferable. None of the five studies made any reference to statistical adjustment for extraneous prognostic factors.

The findings of Werner and Meloy (1992) are perhaps the most striking but also may be the most flawed. Impressively they achieved a 100% accurate prediction that none of those released would violently re-offend, a base rate of zero. Their sample consisted of fifty persons under consideration for release of whom 24 were accepted into a programme for conditional release and it is in these that predictions were deemed to be correct. This provides an extreme example of the difficulty discussed previously (see 1.3.2.) whereby it is not possible to identify false positives because they do not have the opportunity to fail precisely because of this judgement. Without some indication of the rate of false positives perfect predictions lack weight as the 'cut-off' level may lie at too great an extreme. The studies of Sepejak et al. (1983) and Webster et al. (1984) also both show small but significant relationships between clinical predictions of dangerousness and outcome data but

both fail to identify true negatives amongst their outcomes as discussed. Finally the studies of Cocozza and Steadman (1976) and Steadman and Cocozza (1978) found no significant relationship between risk prediction and outcome leading the former authors to conclude 'The data presented constitute the most definitive evidence available on the lack of expertise and accuracy of psychiatric predictions of dangerousness'. However it should be noted that subsequent to the determination of dangerousness being made all defendants remained in either a correctional hospital or a mental hygiene facility for what may have been up to sixteen months prior to release. One hopes and imagines that they received treatment during this period that was aimed at reducing their 'dangerousness' and that should they have been considered to continue to represent a risk they would not have been released. The authors give the impression that all subjects were eventually released implying that they received treatment until this was thought appropriate.

There is thus some evidence that the clinical prediction of dangerousness over two to three years has some limited validity.

1.6.1.3 Predicting recidivism in the long term using clinical judgement

A further five studies addressed the efficacy of clinical predictions of recidivism over the longer term, up to eleven years (Table 3). None of the five studies identified subjects at an early and uniform stage in their criminal histories, three entering patients during there admission for assessment to secure hospital facilities (Quinsey & Maguire, 1996; Kozol et al., 1972; Zeiss et al., 1996), one during their stay in such a facility (Mullen & Reineher, 1982) and the last from those in a programme designed to evaluate the suitability of those already 'shortlisted' for release (Clanon & Jew, 1985). Four studies were subject to bias due to the exclusion of some subjects as a result of their non-release. With Clanon and Jew

Table 3: Studies of Clinical Risk prediction with a follow of more than three years

Authors	number Mean age % Male	Population	Time to follow up	Predictive Method	Outcome Definition	Base Rate	Sens	AUC	True +ve	False +ve	True -ve	False -ve	Correct
Clanon & Jew, 1985	573 N/K 100%	Prison stress assessment unit inmates Predicted dangerous Predicted not dangerous	2-10 years	Positive or negative recommendation for release using Stress Assessment Unit Method- observation in 'stressful' situations	Murder, robbery, assault, rape, other sexual offences	 13% 11%	Chi^2 =NS						
Kozol et al., 1972	435 36 yrs 100%	Pts in a facility for the diagnosis and treatment of dangerous persons	10 years	Clinical interview, psychological tests and historical information	Serious assaultive crimes	11%[a]		0.76[a]	35%[b]	65%[b]	92%[b]	8%[b]	86%[b]
Mullen & Reineher, 1982	165 30 yrs 100%	Forensic psychiatric unit	Mean 2 years	Clinical judgement using psychological tests, rating devices and demographics	Arrest for violent crime	9%		0.58[a] ns	47%[b]	89%[b]	64%[b]	8%[b]	62%[b]

[a] Figures not quoted in paper. Source is Mossman (1994). [b] Figures not quoted in paper. Source is Bjorkly (1995). ns= not significant.
AUC= Area under the ROC curve. Pts= patients. Inpts = Inpatients. Sens = Sensitivity. Spec = Specificity

Table 3: Studies of Clinical Risk prediction with a follow up of more than three years (continued)

Authors	number Mean age % Male	Population	Time to follow up	Predictive Method	Outcome Definition	Base Rate	Other
Quinsey & Maguire, 1996	85 N/K 100%	Maximum security psychiatric institution	11 years	Averaged clinician ratings scored 0-100	4 categories: (1) good (2) economic offences (3) minor offences against person (4) major offences = conviction resulting in sentence of		
	n-28	100%- Psychosis			>5years, or readmission for murder, attempted murder, or wounding.		Correlation 0.07 NS
	n-50	100%-Personality disordered					Correlation 0.05 NS
Zeiss et al. 1996	31 30 yrs 100%	Psychiatric hospital	1-5 yrs	Clinical judgement as	Significant violence.		Chi^2 p+<0,05 Sig
		Predicted violent		'imminently dangerous'		61%	
		Predicted not violent				26%	

[a] Figures not quoted in paper. Source is Mossman (1994). [b] Figures not quoted in paper. Source is Bjorkly (1995). NS= not significant. Sig = Significant. AUC= Area under the ROC curve. Pts= patients. Inpts = Inpatients. Sens = Sensitivity. Spec = Specificity. Corr= Correlation.

(1985) losing up to 11%, Mullen and Reineher (1982) 39%, Kozol et al. (1972) up to 17% and Quinsey and Maguire (1996) retrospectively selecting released patients only. The fifth took all patients for whom extended commitments were initiated (Zeiss et al., 1996) and lists the loss of patients whilst still reporting on their original sample size, the interpretation of which is unclear. The initial referral pattern is likely to have been an additional source of bias in the study of Clanon and Jew (1985) whose subjects were already short listed as potentially fit for release and were further selected as part of the study programme. Quinsey and Maguire (1996) specifically selected patients retrospectively to produce a group with a range of outcomes with which they might best test an actuarial model they were developing in parallel and the remaining two studies fail to provide details of referrals to the study institutions. Both Clanon and Jew (1985) and Quinsey and Maguire (1996) attempted to objectively define their outcome criteria according to certain conviction types with the latter group further categorising them into four outcomes (see Table 3). Kozol et al. (1972) refer only to 'assaultative crimes' and Zeiss et al. (1996) to 'significant violence' whilst Mullen and Reineher (1982) similarly talk only of 'arrest for violent crime'. None of the five studies make any reference to 'blinding' outcome researchers to clinical predictions and only the study of Quinsey et al. (1996) controls for extraneous factors by using a cross-validation sample although this was for the purposes of their actuarial instrument.

Only the studies of Kozol et al. (1972) and Zeiss et al. (1996) found a significant relationship between clinical predictions of violent recidivism and outcome. Kozol et al. (1972) achieved an impressively high 86% of predictions correct. However this must be tempered with the acknowledgement that their study sample were released in batches according to their progress through a treatment programme and only released when thought to be at low risk as evidenced in their 11% base rate recidivism. Zeiss et al. (1996) in their small study which may have lacked complete follow up also found a significant relationship. The remaining

studies failed to find any association between clinical judgement of risk of violent recidivism and outcome.

1.6.2 Studies of actuarial judgement

1.6.2.1 Predicting recidivism within 12 months using actuarial judgement

Literature searches revealed five papers which addressed the efficacy of actuarial predictions of violent recidivism during a follow up period of one year or less (Table 4). As in studies of clinical judgement none identified their cohort at an early and uniform point, one being a retrospective database review (Hedlund et al., 1973), one on emergency room assessments (Gardner et al. 1996), two on 'potentially violent' psychiatric inpatients (Klassen and O'Connor, 1998ab) and one on consecutive emergency commitments (Klassen and O'Connor, 1989). Two of the studies by Klassen and O'Connor (1988a, 1989) lost 9% of participants through refusal or early discharge and in no cases was the referral pattern to the institution adequately described. Three studies achieved complete follow up (Gardner et al., 1996; Klassen and O'Connor, 1989; 1988b) and for one it was not relevant due to the studies retrospective nature (Hedlund et al., 1973). The remaining study lost 6% to incomplete data (Klassen and O'Connor, 1988b). For the three studies by Klassen and O'Connor (1988ab; 1989) outcomes were defined in terms of the crimes considered violent. Gardner et al. (1996) used an eight point outcome scale whilst that of Hedlund et al. (1973) was least well defined being according to patient self report when asked unspecified items on an interview checklist. Only two studies had attempted to blind outcome researchers to predictive status (Klassen and O'Connor, 1989; 1988b). Despite the greater need in actuarial assessment to cross validate results on a sample other than that on which a measure was developed this was only conducted by two studies (Hedlund et al., 1973; Klassen and O'Connor, 1989) and acknowledged as desirable but impractical due to sample size by the third (Klassen and

Table 4: Studies of Actuarial Risk prediction with a follow up of less than one year

Authors	number Mean age % Male	Population	Time to follow up	Predictive Method	Outcome Definition	Base Rate	Sens	AUC	False +ve	False -ve	Correct
Gardner et al., 1996	784 29 yrs N/K	Psychiatric emergency room pts 31%-Substance abuse 20% Affective disorders 19%-Personality disorders	4-6 months	Complete screen Historical variables Current variables	Seriousness of violence reported in follow up interview			0.74 0.71 0.70	42%		
Hedlund et al., 1973	2,762 43 yrs 60%	Psychiatric inpts. 27%-Substance abuse 17%-Organic brain syndromes 16%-Schizophrenia	1 month	Discriminant analysis, prospectively tested	Assaultive behaviour	10%[a]		0.76[a]	60%	8%	89%
Klassen & O'Connor, 1988a	239 N/K 100%	Pts from a community mental health centre for short term emergency commitments 45%-Schizophrenia 26%-Substance abuse 8%-Organic brain syndrome	6 months (at liberty for at least 4 months)	Stepwise discriminant analysis	Arrested or re-admitted for violence. Noted from arrest and mental health centre records	19%[a]	76%	0.91[a]	41%	6%	85%

[a] Figures not quoted in paper. Source is Mossman (1994). [b] Figures not quoted in paper. Source is Bjorkly (1995). AUC= Area under the ROC curve.
Pts= patients. Inpts = Inpatients. Sjs = subjects.

Table 4: Studies of Actuarial Risk prediction with a follow up of less than one year (continued)

Authors	number Mean age % Male	Population	Time to follow up	Predictive Method	Outcome Definition	Base Rate	AUC
Klassen & O'Connor, 1988b	109 N/K 100%	Short term psychiatric ward inpts 100%-Schizophrenic	1 year (at liberty for at least 3 months)	Discriminant analysis	Arrest or re-admission due to violence	32%[a]	0.94[a]
	127 N/K 100%	100%-Non-schizophrenic				28%[a]	0.98[a]
Klassen & O'Connor, 1989	265 N/K 100%	Pts from a community mental health centre for short term emergency commitments	1 year (at liberty for at least 3 months)	Formula derived from multiple regression	Arrest or re-admission due to violence	25%[a]	0.76[a]

[a] Figures not quoted in paper. Source is Mossman (1994). [b] Figures not quoted in paper. Source is Bjorkly (1995). AUC= Area under the ROC curve.
Pts= patients. Inpts = Inpatients. Sjs = subjects.

O'Connor, 1988a). The remaining study by Klassen and O'Connor (1988b) does not address the issue and that by Gardner et al. (1996) statistically adjusts for differences between its sample and the population from which it was drawn but fails to apply its equation to a new sample.

The studies achieve remarkably high values for their AUC's ranging from 0.7 to 0.98. The samples in which cross validation samples were employed, and which should therefore be regarded as more accurate both achieved, coincidentally, AUC's of 0.76 (Klassen and O'Connor 1989; Hedlund et al., 1973). These results suggest that actuarial measures can predict violent recidivism over a 12 month period at levels significantly greater than chance.

1.6.2.2 Predicting recidivism in one to three years using actuarial judgement

Five studies predicting violent recidivism during a one to three year period using actuarial data were identified (Table 5). The samples did not select patients at an early and uniform stage in their careers instead taking, retrospective hospital discharges (Cocozza and Steadman, 1974; Coccozza et al., 1978), patients at pre-trial assessment (Sepejak et al., 1983; Menzies et al., 1985) or on release from prison (Holland et al. 1982). In the two retrospective studies all subjects were selected by virtue of having been released. In two studies all patients were followed up regardless of destination (Sepejak et al., 1983; Menzies et al , 1985; although in the latter sample only one of 203 was not released) but in the study of Holland et al. (1982) 42% were not released and therefore excluded from follow up. None of the authors described the referral patterns to their study institutions in detail although all managed to account for all subjects at follow up. The objectivity of outcome criteria varied from good, its categorisation on eleven point scales (Sepejak et al., 1983; Menzies et al., 1985), to categorised crimes (Cocozza et al. 1978), and violence against

Table 5: Studies of Actuarial Risk prediction with a follow up of two to three years

Authors	number Mean age % Male	Population	Time to follow up	Predictive Method	Outcome Definition	Base Rate	Corr	AUC
Cocozza & Steadman, 1974	98 N/K 100%	Psychiatric pts	Mean 28.5 months	Legal Dangerousness Scale and age, retrospectively	Violent assaults	20%[a]		0.89[a]
Cocozza et al. 1978	1,920 49 yrs 56%	Psychiatric pts	Mean - 19 months	Number of total prior arrests	Violent crime	0.9%	0.11	0.85[a]
	1,938 App. 41yrs 56%					1.7%	0.21	0.93[a]
Holland et al. 1982	198 28 yrs 100%	Prisoners	32 months	Past violent convictions		11%[a]		0.54[a]
				Past non-violent convictions		11%[a]		0.61[a]

[a] Figures not quoted in paper. Source is Mossman (1994). [b] Figures not quoted in paper. Source is Bjorkly (1995). AUC= Area under the ROC curve.
Pts= patients. Inpts = Inpatients. Sjs = subjects. Corr= Correlation.

Table 5: Studies of Actuarial Risk prediction with a follow up of two to three years (continued)

Authors	number Mean age % Male	Population	Time to follow up	Predictive Method	Outcome Definition	Base Rate	AUC	True +ve	False +ve	True -ve	False -ve	Correct
Menzies et al., 1985	203 62%<30yrs 89%	Pre-trial forensic clinic pts	Mean 17.3 months	Discriminant analysis to develop Dangerous Behaviour Rating Scheme (DBRS)	Dangerous behaviours noted in criminal charges, prison & hospital files	29%[a]	0.67[a]	82%[b]	39%[b]	58%[b]	19%[b]	69%[b]
Sepejak et al., 1983	360 app.16-30yrs 'majority'	Forensic brief assessment unit pts	2 years	Previous incarceration	Criminal & institutional behaviour rated on 11 point scale	47%[a]	0.59[a]					

[a] Figures not quoted in paper. Source is Mossman (1994). [b] Figures not quoted in paper. Source is Bjorkly (1995). AUC= Area under the ROC curve.
Pts= patients. Inpts = Inpatients. Sjs = subjects. Corr= Correlation.

persons (Cocozza and Steadman, 1974) to conviction for threatened or actual physical injury (Holland et al., 1982). Only two studies made reference to assessor blinding with one employing independent raters (Sepejak et al., 1983) and one employing independent judges to place subjects on a rating scale according to a profile but it does not appear that the researcher compiling the profile was blind to predicted outcome (Menzies et al., 1985).

All five studies reported significant success at predicting violent recidivism by actuarial means. The strongest predictor was a combination of the Legal Dangerousness Scale and age with an AUC of 0.9 although this was in a small sample that did not undergo cross-validation.

1.6.2.3 Predicting recidivism in the long term using actuarial judgement

Seven studies were identified examining the actuarial prediction of risk for follow up periods of longer than three years (Table 6). Six failed to achieve an inception cohort at an early and uniform point in patient careers recruiting them retrospectively (Steadman & Morrissey, 1982; Black and Spinks, 1985), on release from prison (Serin and Amos, 1995), a secure hospital (Harris, 1991), on admission to a pre-trial forensic clinic (Menzies, 1994) or a secure hospital for assessment (Harris, 1993). The seventh study was an excellent example of a cohort study recruiting eight year olds on a geographical basis for long term follow up (Farrington, 1989). Two studies lost subjects to non-release (Steadman & Morrissey, 1982; Harris, 1993). Steadman and Morrissey (1982) did so substantially, losing 40%, 20% and 53% in their three samples although it is unclear whether some losses were due to an inability to trace subjects. Eleven percent of the original sample of Harris et al. (1993) were not released. All studies but that of Farrington (1989) and Harris (1991; 1993) provided a poor level of detail regarding referral criteria to their institution. Significant numbers were

Table 6: Studies of Actuarial Risk prediction with a follow up of 3 years or more

Authors	number Mean age % Male	Population	Time to follow up	Predictive Method	Outcome Definition	Base Rate	Other	AUC	False +ve	Correct
Black & Spinks, 1985	125 41 yrs 100%	Special hospital pts	5 yrs	Offence & age at discharge, MMPI scale, F, Ex.	Re-conviction	10%		0.94		
Farrington, 1989	411 32 yrs 100%	Age & geographically defined cohort	15 yrs	Formulae derived from multiple regression	Conviction for violence	12%[a]		0.86[a]		
Harris et al., 1993	618 N/K 100%	Referrals to secure hospital	81 months (Almost 7 years)	VRAG	Charged or re-admitted for violent offence	31%	Sens= 60% Spec= 78%		45%	73%
Harris et al., 1991	169 <25yrs 100%	Pts released from a secure hospital therapeutic programme	Mean 4yrs Up to 10 yrs	PCL-R >25	Violent offence					
		All patients				40%				76%
		Psychopaths				70%				78%

[a] Figures not quoted in paper. Source is Mossman (1994). [b] Figures not quoted in paper. Source is Bjorkly (1995). AUC= Area under the ROC curve.
Pts= patients. Inpts = Inpatients. Sjs = subjects. Corr= Correlation.

Table 6: Studies of Actuarial Risk prediction with a follow up of 3 years or more (continued)

Authors	number Mean age % Male	Population	Time to follow up	Predictive Method	Outcome Definition	Base Rate	Corr	AUC	False +ve	False -ve	Correct
Menzies et al., 1994	136 60% <30yrs 86%	Pre-trial forensic clinic pts	6 yrs	Dangerous Behaviour Rating Scale	Violent transactions		0.15 sig				
Serin & Amos, 1995	300 31 yrs 100%	Prisoners	Mean 5.5 yrs	Hare Psychopathy Checklist >29	Conviction for violent offence	17%					80%
Steadman & Morrissey, 1982	154 app. M-31 yrs 100%	Unfit to stand trial for felony	Several mths to yrs	Previous arrest for violent crimes & age at 1st hospitalisation	Physical attack on another	18%			50%	15%	82%
	(n=257)	(inc unreleased sjs)				(39%)[a]		(0.63)[a]			
	227 app. M-31 100%	Charged with felony but not indicted	"	"	"	20%			100%	18%	83%
	(n=282)	(inc unreleased sjs)				(30%)[a]		(0.57)[a]			
	117 app. M-35 100%	Psychiatric admissions	"	"	"	19%			100%	19%	80%
	(n=250)	(inc unreleased sjs)				(8%)[a]		(0.59)[a]			

[a] Figures not quoted in paper. Source is Mossman (1994). [b] Figures not quoted in paper. Source is Bjorkly (1995). AUC= Area under the ROC curve.
Pts= patients. Inpts = Inpatients. Sjs = subjects. Corr= Correlation.

lost to follow up by only two studies. Farrington (1989) failed to follow up 6% of their patients but given the length of their study this does not appear unreasonable. Menzies et al. (1994) 'selects' 162 subjects from their previously reported (see Table 5) sample of 203 but it is not clear by what criteria or if this is the result of default due to 'lost' subjects. Outcome criteria were ill defined including physical attacks on another (Steadman and Morrissey, 1982), a violent offense (Harris, 1991; 1993), assault (Black and Spinks, 1985), reconviction for a violent offence (Serin and Amos, 1995), convictions for violence (Farrington, 1989) and violent transactions defined according to a coding manual the details of which are unavailable (Menzies et al., 1994). None of the outcome researchers for any study were blinded to predictive status.

All seven studies succeeded in predicting violent recidivism at levels above chance although for two of these the results were borderline (Steadman and Morrissey, 1982; Menzies et al., 1994). The remaining studies, inclusive of that which appeared most methodologically valid (Farrington et al., 1989) each predicted violent recidivism at significantly above chance levels, four using their own actuarial models (Farrington et al., 1989; Black and Spinks, 1985; Harris 1991; 1993) and one the Hare Psychopathy Checklist (Serin and Amos, 1995).

1.6.3 Literature Review Conclusions

The current review of the literature identified twelve studies of clinical judgement and seventeen of actuarial with few trends in their results being evident across time frames. Of the studies of clinical judgement only four showed a significant relationship between prediction of violent recidivism and outcome, although a further two did achieve borderline significance. Of the studies of actuarial judgement all illustrated a significant relationship between actuarial prediction of violent recidivism and outcome although two of these only

reached borderline significance. One study contributed results to both our discussion of clinical and actuarial instruments comparing their use in the same sample (Gardner et al., 1996). They concluded 'Actuarial prediction of patients' violence was substantially more accurate than clinical prediction. The actuarial screens had more favourable trade-offs between false-positive and false-negative errors than did predictions based on clinicians' concerns about patient violence' (Gardner et al., 1996). None of these studies are methodologically flawless and this review is not exhaustive, but it does indicate a clear trend in favour of actuarial methods. Meehl (1997) perhaps stated the case against clinical judgement most strongly:-

> "Since clinical experience consists of anecdotal impressions by practitioners, it is unavoidably a mixture of truths, half-truths, and falsehoods. The scientific method is the only known way to distinguish these, and it is both unscholarly and unethical for psychologists who deal with other persons' health, careers, money, freedom, and even life itself to pretend that clinical experience suffices and that quantitative research on diagnostic and therapeutic procedures is not needed."

1.6.4 Two promising instruments

Only two instruments have undergone substantial development and validation work whose progress appears promising, the HCR-20 (Webster et al., 1997) and the VRAG (Quinsey et al., 1998).

The HCR-20 is a risk assessment instrument comprising ten historical items, five clinical items reflecting dynamic correlates of violence and five risk management items addressing factors which may increase or decrease risk (Douglas, 1999). The inclusion of dynamic factors is an attempt to bridge the gap between clinical and actuarial instruments as the latter have been criticised for the inability of patients to improve their scores over time. A number of studies have been conducted addressing its reliability but unfortunately all have yet to be published, hence there omission from our literature review. The results to date are promising and summarised in the table below.

Table 7: Summary of HCR-20 Studies

Authors	n	Follow up period	Outcome	AUC	Comment
Douglas et al, 1998	193	2 years	Violent crime	0.80	PCL-R AUC = 0.79
Ross et al, 1998	101	time unclear	Violent crime	0.75	
Grann et al 1998	404	2 years	Violent reconviction	0.71	Historical scale only, VRAG AUC =0.68
Strand et al 1998	40	3-12 years	Violence	0.80	

PCL-R = Psychopathy Checklist Revised (ref)

Equally promising is the VRAG which forms the basis of the present study.

1.7 The Violence Risk Appraisal Guide.

The VRAG was developed at Penetanguishene in Canada, a maximum security mental health facility which accepts criminal offenders and mentally ill men with a history of violence. The study sample consisted of 332 men admitted to the facility and a comparison group of 286 who were assessed there. Each man in the second group was matched with one in the first on the basis of index offence, age at index offence, previous violent and non-violent criminal acts and the time of the index offence. For each subject information on a broad range of variables was collected from case files, themselves compiled from a number of external sources, by trained research assistants. Outcome data was collected from a range of criminal record sources by trained research assistants other than those involved in the initial data collection from hospital files. The sample of 618 men were selected from an original group of 685 by virtue of having the opportunity to fail defined as release to an open psychiatric ward, a halfway house or into the community. Failure was defined by the commission of a violent act from common assault to murder and included sex offences, armed robbery, forcible confinement, threatening with or pointing a firearm or admission to a secure facility for violence towards persons which would normally result in criminal charges. On average the men met criteria for eligibility to fail for a period of 81.5 months (almost seven years) prior to follow up. During this time 191 (31%) met the criteria for violent failure, although as with all such follow up studies base rate recidivism is likely to be an underestimate (Webster et al., 1994).

Initial variable selection was based on the published prediction literature and hypotheses on theoretical grounds producing 50 variables in the areas of demographics, childhood history, adult adjustment, index offence, criminal and psychiatric history and assessment results which were subject to separate stepwise discriminant analyses to select those for inclusion in the VRAG (Webster et al., 1994; Rice, 1997). The sample was then split in two, first

randomly and then according to whether disposal was to prison or hospital. For both the entire sample and each of the subgroups multiple regression analyses were used to identify up to four variables for inclusion in a stepwise regression. At analysis completion 12 variables remained to form the VRAG (Rice, 1997).

Of the twelve VRAG variables nine are historical and should be available in institutional records. These include childhood variables; whether separated from either parent prior to age 16 and elementary school maladjustment score; Index offence variables; whether there was a female victim, victim injury and age at time of index offence; Criminal history; non-violent offence history score and failure on prior conditional release; history of alcohol abuse in the offender, their family and in offence history and whether the subject was ever married. The remaining three variables are whether the patient meets DSM-III criteria for any personality disorder or schizophrenia, which one would hope would be in medical records, and their PCL-R score. Thus eleven of the VRAG's variables can be collected by a non-clinically qualified rater from official records.

During assessment the VRAG's twelve items are assigned raw scores (see Appendix 1) which are then translated to a scale score of between one and nine representing nine equal range (of 8 raw score points) raw score categories with a higher score reflecting an increased risk of recidivism. Analysis illustrated a positive correlation between increasing VRAG scores and likelihood of violent recidivism with an AUC of 0.76. Analysis also found the VRAG score to be positively correlated with violent re-offending, to the severity of those offences and to how quickly, after release from maximum security, they occurred (Quinsey et al., 1998). Examination found accuracy to remain high even when the base rate was artificially altered by changing the definition for failure or time to follow up (Rice and Harris, 1995).

The nature of its development had certain strengths that are likely to increase the VRAG's reliability and validity. It was developed on a large sample, over a long follow up period, using comprehensive and well documented file information, including criminal justice and clinical data, collected by trained data collectors and analysed using powerful statistical methods (Webster et al., 1994). Inter-rater reliability on a sample of 20 achieved a correlation of 0.9 (Harris et al., 1993) and standard error was equal to only half a scaled score range (4.1) (Harris et al., 1993). Validation studies extending follow up to ten years, adding sex offenders and studying only sex offenders continued to yield ROC AUC's of 0.73 to 0.77 (Quinsey et al., 1995). Cross validation has been undertaken in 159 sex offenders achieving an ROC AUC of 0.77 (Rice and Harris, 1997), in 59 sex offenders (using a modified version) (Belanger & Earls, in press) and in predicting institutional misconducts (Kroner and Mills, 1997).

1.7.1 The Psychopathy Checklist -Revised and the VRAG

The most powerful predictor of the twelve variables that comprise the VRAG is the Psychopathy Checklist Revised (PCL-R) Score correlating with violent recidivism at 0.34 (Webster et al., 1994). However, as previously noted, it is also the one variable that is not routinely available in official records precluding the scoring of the VRAG from records alone. The PCL-R is a twenty item taxon scale designed to indicate the likelihood of a psychopathic personality disorder. Each item is scored on a three point scale of '0' 'does not apply', '1' 'uncertain, applies to a certain extent' or '2' 'applies' giving a score range of 0-40 where a score over 30 or above is considered diagnostic (Hart et al., 1994). The scale has two factors, the first reflecting interpersonal and affective characteristics such as egocentricity, manipulativeness, lack of remorse and callousness (Hart et al., 1994). The second identifying trends towards impulsivity, an anti-social and unstable lifestyle or social

deviance (Hart et al., 1994). Scoring of the PCL-R requires a thorough examination of official records followed by a semi-structured interview of between one and a half and two hours resulting in a full assessment time of between two and three hours (Hart et al., 1994). However it can, should records be of a high enough quality, be scored by case note review alone (Hare, 1991). Not only does the PCL-R require an extensive use of clinical time but in addition requires that clinicians be:-

- In possession of a degree in social, medical or behavioural sciences, such as a Ph.D, D.Ed. or M.D.

- Be registered with the local state or provincial registration body that regulates the assessment and diagnosis of mental disorder.

- Have experience with forensic populations (as demonstrated by registration as a diplomate in forensic psychology or psychiatry, completion of a practicum or internship in a clinical-forensic setting, or at least two years of relevant work-related experience).

- Ensure they have adequate training and experience in the use of the PCL-R.

(Hare, 1991)

Thus the VRAG pays a very high price in administration time and expertise to its best predictive variable, which when added to scoring the remaining 11 variables, reduces the VRAG's clinical utility.

Quinsey et al. (1988) hypothesised that for the purposes of the VRAG it may be possible to differentiate psychopaths and non-psychopaths using fewer, more easily rated items than contained in the PCL-R. To this end they constructed an eight item scale of historical items that could be ascertained from clinical records, the Childhood and Adolescent Taxon Scale (CATS). In subsequent analysis the replacement of the PCL-R with the CATS resulted in almost identical predictive validity in the VRAG, an AUC of 0.75. A cross-validation study on 54 Canadian mentally disordered offenders found the correlation between VRAG scores calculated using the PCL-R and the CATS to be 0.975 (Quinsey et al., 1997).

Should this finding prove reliable, the implications of this work will be profound for the field of risk prediction. The recent self doubt created by studies of clinical predictive validity prompting debate regarding whether we should even attempt risk prediction could be put aside with the adoption of a short records based assessment that can be scored by non-clinical raters to provide prediction of risk of violent recidivism. This may contribute to meeting the increased demand for time efficient and reliable risk assessment measures which have resulted from various community and criminological factors and which may increase further if the recommendations of recent reports are implemented (MacClean, 1999; Fallon et al., 1999; Reed, 1994).

The following study addresses the clinical questions that must be answered before the VRAG, calculated using the CATS rather than the PCL-R, can be widely utilised.

1.8 Statement of Hypotheses

In a UK inpatient forensic population:-

1) The CATS is a valid alternative to the PCL-R in calculating the VRAG.

   Both in :

   a) Comparison to interview based PCL-Rs.

   b) Comparison to records based PCL-Rs.

2) It will have good inter-rater reliability.

Further Questions to be examined

1) What is the nature of the difference in VRAG scores calculated using the CATS and

   a) Interview based PCL-Rs?

   b) Records based PCL-Rs?

2) What is the extent of measurement overlap between the CATS and PCL-R items?

2 Methods

2.1 Subjects

The study sample comprised 120 subjects who were presently, or had been, admitted to the State Hospital, Carstairs. This is a maximum security facility which accepts patients 'from hospitals, the courts and from prisons because of mental illness and their dangerous, violent or criminal propensities' (The State Hospitals Board for Scotland, 1999). Historically the hospital performed a more custodial role reflecting its management by the prison service but since 1991, when it became part of the National Health Service, its role has been therapeutic and similar to that performed by the Special Hospital system in England.

"The Hospital caters for up to 245 patients from Scotland and Northern Ireland who are admitted under the requirements of the Mental Health (Scotland) Act 1984 and related legislation because of their dangerous, violent or criminal propensities. Patients fall into two categories: State patient, i.e. those admitted from the Courts and Prisons under a Restriction Order and who may be discharged only with the approval of the Secretary of State for Scotland, and Non-State patients, i.e. those admitted either from the Courts but without a Restriction Order or as ordinary State Hospital patients usually on transfer from NHS hospitals, who may be discharged by the State Hospital doctor in charge with the agreement of the State Hospital Managers." (The State Hospitals Board for Scotland, 1998).

As the study required that a trained rater had administered the PCL-R on all subjects they were selected as a convenience sample as participants of two previous studies.

Study 1: As part of a study on the inter-rater reliability of the PCL-R 58 complete PCL-Rs were available on current patients that had been administered by one trained administrator

(JD) using both a structured interview and case note review, hereafter referred to as 'interview based' PCL-Rs.

Study 2: As part of a cohort study of patient characteristics the PCL-R had been calculated on 171 subjects from case note review only, hereafter referred to as 'records based' PCL-Rs. These patients were those who comprised a sample who had been discharged from the hospital between 1985 and 1987. Of these a convenience sample of 62 were taken in numerical order but partially selected to avoid both females and the learning disabled to maintain sample homogeneity. Prior to a decision on the inclusion of females being taken one was included in one inter-rater reliability sample (BH). The decision to 'select out' some learning disabled subjects was taken when it became apparent that the second sample was severely skewed in this direction reflecting the then population of the hospital. This was to allow the sample to better reflect the current hospital population, to allow our results to be most applicable to primarily mentally disordered populations and to increase its comparability with study sample one.

2.2 Data Collection

2.2.1 Computing the VRAG

VRAGs and CATS were completed on all patients in the recommended manner by case note review. The State Hospital maintains extensive case notes in addition to holding former hospital and prison records although it was notable that those in the second sample, being 12 to 14 years older that those of sample one, were of a lower quality.

Prior to data collection, scoring criteria for each VRAG and CATS variable were agreed to further clarify those already provided by the development group (Quinsey et al., 1998) (Appendix 3).

For the purposes of the present study 125 VRAGs were completed comprising 58 from study one, interview based PCL-Rs and 67 from study 2, records based PCL-Rs, by three separate raters who completed 100 (SOR) inclusive of all subjects in the interview based PCL-R sample, 12 (BH) and 13 (KR) respectively.

2.2.1.1 Sample homogeneity

Ideally we wished to combine all 125 VRAGs into one data set but we hypothesized that interview based PCL-Rs and records based PCL-Rs may differ as might rater reliability. Thus the first two stages of analysis were to examine the validity of the CATS as an alternative to the PCL-R in calculating the VRAG separately for interview and records based PCL-Rs and separately by raters. It quickly became apparent that there were indeed some differences both between interview and records based PCL-Rs and between raters. The decision was taken therefore to continue to conduct all analyses (with the exception of factor analysis) separately for the two alternative means of computing the PCL-R and, having examined the difference between raters, to remove data from the two additional raters from the primary analyses.

2.3 Analyses

2.3.1 Is the CATS a valid alternative to the PCL-R in calculating the VRAG?

In comparison to interview based PCL-Rs?

In comparison to records based PCL-Rs?

Analyses first addressed our primary hypothesis, whether the CATS was a valid alternative to the PCL-R, in calculating the VRAG using the CATS first for subjects with interview based PCL-Rs and for those with records based PCL-Rs. This hypothesis was analysed using both a Pearson correlation and by plotting scores on a scatterplot inclusive of a regression trend line and 95% confidence intervals.

2.3.2 Are these results reliable or simply the result of one good rater?

We next addressed the influence of raters on the validity of the CATS as an alternative to the PCL-R in VRAG computation. Again, we analysed the impact of replacing the PCL-R with the CATS but this time for each rater, using Pearson correlations and scatterplots with regression trend lines and 95% confidence intervals. Due to the small sample sizes with resulting restriction of variance it became necessary to adjust the correlations for the samples of the additional raters BH and KR to compensate for the restriction in range of the measurement in one or both variables (Glass & Hopkins, 1996).

2.3.3 What is the nature of the difference in VRAG scores calculated using the CATS and the PCL-R?

The nature of the difference in VRAG scores calculated using the CATS as an alternative to the PCL-R was examined separately for the three samples in turn, interview based PCL-Rs, records based PCL-Rs and for the combined sample.

It was addressed in four ways:

(i) What is the difference in VRAG scaled scores calculated using the CATS rather than the PCL-R?

For each sample the frequency of each value of error score was calculated and plotted on a histogram allowing description of the size of errors and their direction.

(ii) Is the validity of the CATS as a replacement for the PCL-R maintained across VRAG score ranges?

VRAG scaled score ranges for each sample were grouped into scores 1-5 and 6-9 allowing examination using a Pearson correlation of the validity of the CATS as a replacement for the PCL-R in the VRAG at each end of the VRAG scale. As the variance of VRAG scaled scores in the 6-9 range was restricted as a result of a small sample size the correlations were again dis-attenuated to compensate for this.

(iii) The accuracy of predicted VRAG scaled scores using the CATS in comparison to the PCL-R.

Regression analyses were conducted to produce predicted values using the CATS for each VRAG scaled score traditionally calculated using the PCL-R in addition to both the point

and mean value confidence intervals. Both the predicted score and confidence intervals were illustrated in a line plot.

(iv) The comparative distribution of total VRAG scores calculated using the CATS in comparison to the PCL-R.

For each sample the comparative distribution of total VRAG scores calculated using the CATS vs. the PCL-R was illustrated graphically and examined for a significant difference in distribution using the Chi Square goodness of fit statistic.

2.3.4 VRAG variable scores for subjects in the interview based, records based and combined PCL-R sample

For descriptive purposes the distribution of each VRAG variable for both the records and interview based PCL-Rs and the combined sample was calculated (Appendix 4).

2.3.5 How does the CATS replicate PCL-R scores, in what way are their composite items related?

The following analyses were each conducted on three samples, the records based PCL-Rs, interview based PCL-Rs and the whole sample where considered relevant in order to examine any influence alternative means of assessing the PCL-R may have on score distribution and inter-relations.

(i) The distribution of individual CATS & PCL-R variables

To better understand how the CATS replicates the PCL-R we first examined the basic frequency of distribution for each CATS variable (Appendix 5) and each PCL-R variable (Appendix 6) for the three samples.

(ii) The inter-correlation of CATS variables

The internal structure of the CATS scale was examined by constructing a correlational table of each of the CATS variables against each other for the combined sample.

(iii) A comparison of the distribution of total VRAG scaled scores for interview and records based PCL-Rs.

The distribution of each sample's scaled scores were compared by adjusting the records based scores (n=42) to allow for the difference in sample size in comparison to the interview based sample (n=58). Due to the small numbers in some groups scaled scores were grouped in the ranges 1-3, 4-6 and 7-9. The distribution of expected versus observed frequencies was examined using the Chi Square statistic.

(iv) Correlational analysis of the relationship between CATS and PCL-R variables

The relationship between each CATS variable and each PCL-R variable and its factors was examined for the combined sample by again constructing a correlational table comparing all variables.

(v) Factor analysis of combined CATS and PCL-R variables

To further clarify the inter-relationship between CATS and PCL-R items all variables, eight from the CATS and twenty from the PCL-R, were included in a factor analysis of the combined sample (n=100). To allow tests of factor extraction quantity we began with exploratory Principal Component and Image Component Analyses. The latter also allowing

the precise estimation of common variance available for extraction and providing the same methods of factor extraction quantity that were used as 'replication' indices under the conditions of specific variance removal. The PsWin FACTOR program of Barrett (1996) was used for these calculations.

Three tests of factor extraction quantity were used to estimate the number of factors amongst the combined CATS and PCL-R items. These were the Velicer Minimum Average Partial (MAP) Test (Velicer, 1976) which derives the number of likely factors from analysis of the sum of partial correlations. Armor's theta (Armor, 1974), which is similar to a Cronbach coefficient alpha but calculated on an unrotated component factor, and Barrett and Kline's autoscree (1982), a computational form of Cattell's (1966) manual scree test which attempts to identify the 'scree' break-point as a slope discontinuity within a series of data points.

These analyses were followed by the computation of a Minimum Residual (MINRES) Factor Analysis which requires that the number of factors be specified in advance, in contrast to Principal Components Analysis and Image Component Factor Analysis. Rather than extracting common factors that maximise the variance to be extracted by each factor, MINRES extracts factors that maximally reproduce the observed variable correlations. Since this criterion more closely matches the fundamental goal of factor analysis, MINRES is thus to be considered a more appropriate factor extraction technique.

The resulting factors were then rotated from maximum orthogonality to maximum obliquity using the Barrett and Kline (1982) modified Jennrich and Sampson (1966) direct oblimin procedure. In keeping with the goal of 'simple structure' the optimal solution chosen is one where the +-0.1 hyperplane count is at its maximum. The PsWin ROTATE program was used for this procedure.

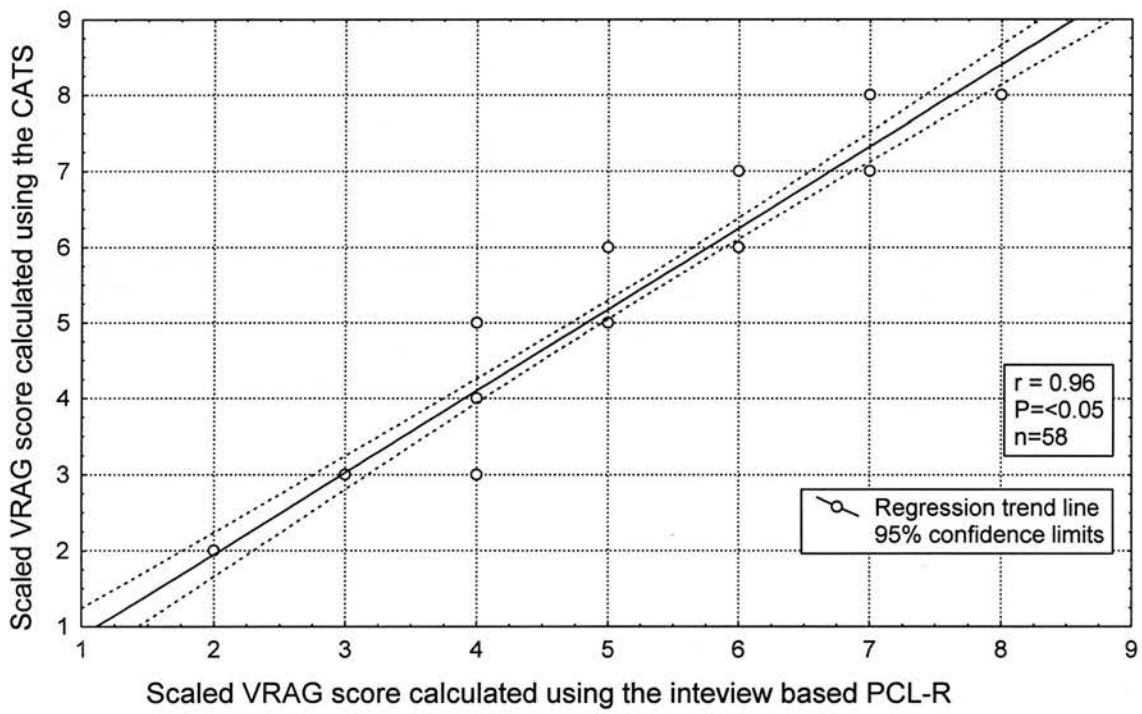
3 Results

3.1 Is the CATS a valid alternative to the PCL-R in calculating the VRAG?

3.1.1 In comparison to interview based PCL-Rs?

The Pearson correlation of VRAG scores calculated using the interview based PCL-R score in comparison to those using the CATS score was 0.96 (P=<0.05; n=58). The distribution of scores, the regression trend line and 95% confidence intervals are plotted in Figure 1.

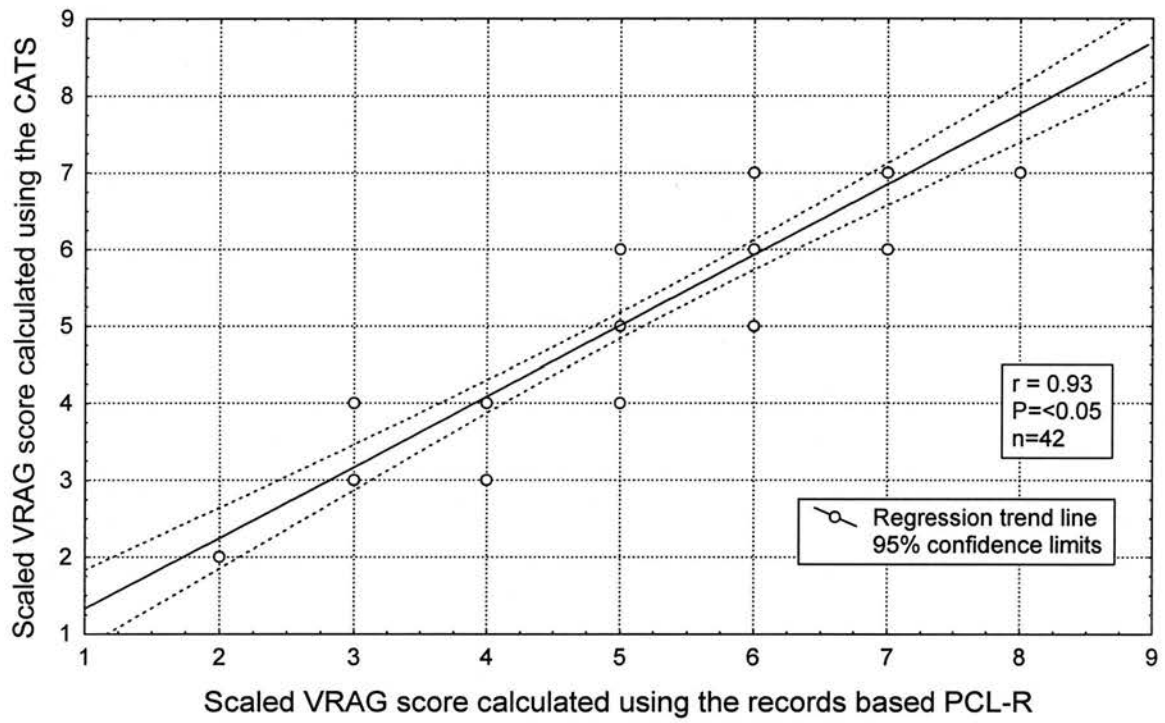
Figure 1: Scatterplot of the correlation of VRAG scores calculated using either an interview based PCL-R or the CATS with regression trend line and 95% confidence intervals.



3.1.2 In comparison to records based PCL-Rs?

Comparison of VRAG scores calculated using records based PCL-Rs or the CATS using Pearson correlation gave a value of 0.93 (P=<0.05; n=42). Again, these are illustrated in scatterplot form with a regression trend line and 95% confidence intervals (Figure 2).

Figure 2: Scatterplot of the correlation of VRAG scores calculated using either records based PCL-R or the CATS with regression trend line and 95% confidence intervals.



3.2 Are these results reliable or simply the result of one good rater?

The two additional raters had calculated VRAG scores on records based PCL-Rs only. The Pearson correlation between their VRAG scores using records based PCL-Rs vs. the CATS was 0.88 (Rater BH; P=<0.05; n=12) and 0.81 (Rater = KR; P=<0.05; n=13) (Table 8). A scatterplot illustration of their score distribution is illustrated in Figure 3 (BH) and Figure 4 (KR).

Figure 3: Scatterplot of the correlation of VRAG scores calculated by BH using either a records based PCL-R or the CATS with regression trend line and 95% confidence intervals.

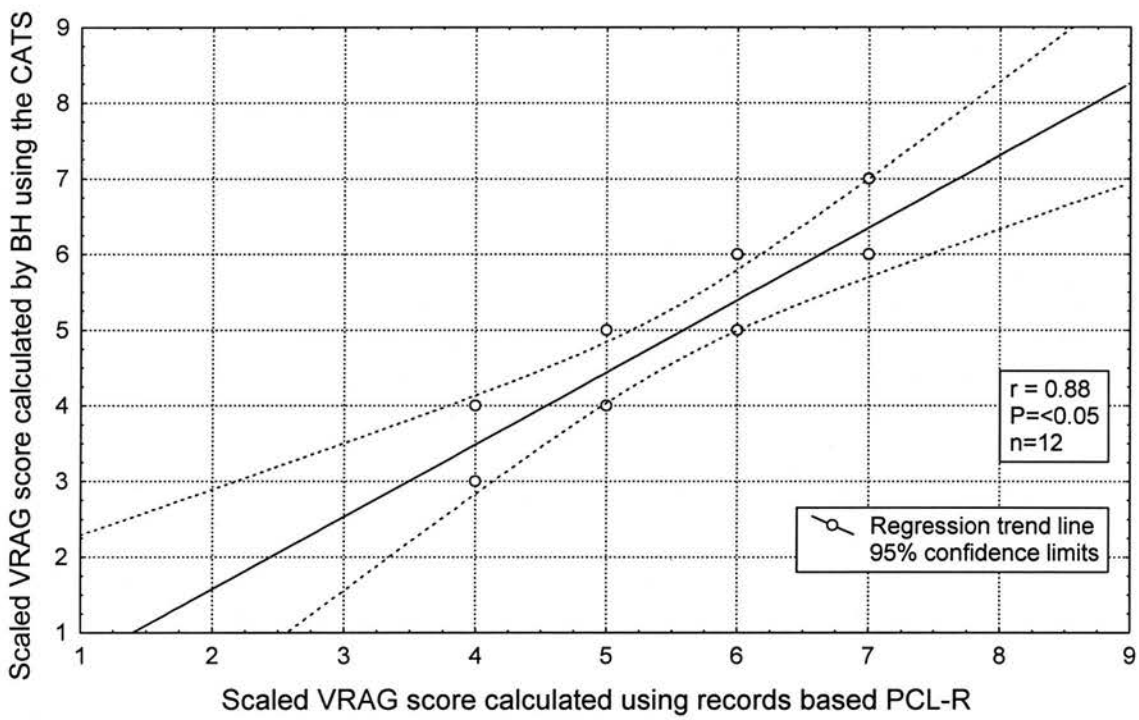


Figure 4: Scatterplot of the correlation of VRAG scores calculated by KR using either a records based PCL-R or the CATS with regression trend line and 95% confidence intervals.

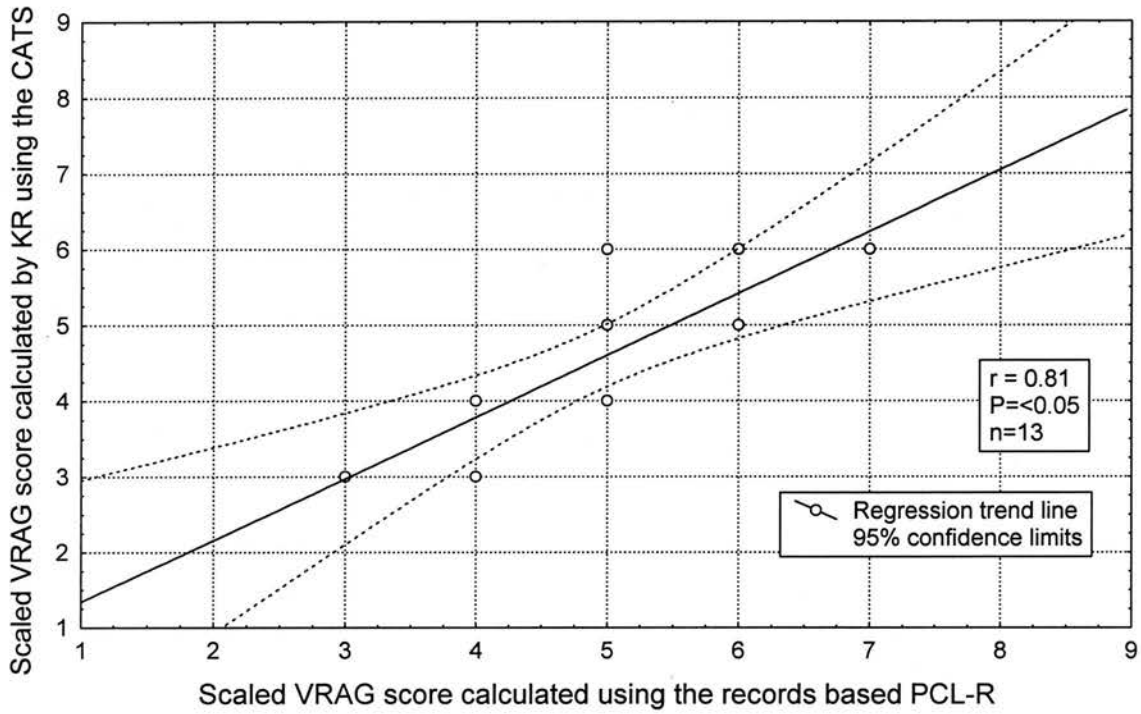


Table 8: Pearson Correlation between VRAGs calculated by each rater using either interview or records based PCL-Rs versus the CATS.

Rater	Interview based PCL-Rs						Records based PCL-Rs					
	n	SD		Correlation			n	SD		Correlation		
		PCL-R	CATS	Raw	Sca	Adj		PCL-R	CATS	Raw	Sca	Adj
SOR	58	1.51	1.66	0.51	0.96	*	42	1.47	1.45	0.27	0.93	0.95
BH	-	-	-	-	-	-	12	1.00	1.08	0.37	0.88	0.97
KR	-	-	-	-	-	-	13	1.04	1.05	-0.29	0.81	0.95

Sca= Scaled VRAG score, correlation between scaled VRAG score calculated using the PCL-R vs. the CATS. Adj= The adjusted correlation taking into account homogeneity of variance. *Not calculated as this was used as the 'population' SD.

On initial observation results appear to indicate an effect of rater upon the reliability of the CATS as a substitute for the PCL-R in the VRAG, however it is noteworthy that the small sample sizes for the latter two raters (BH; KR) have led to increased homogeneity of variance as reflected in their lower standard deviations. As there is a monotonic relationship

between SD attenuation and correlation size this may, in part, explain the difference in correlation size between raters. We therefore double corrected, for both restriction in the CATS and the PCL-R, to create a dis-attenuated correlation for each group. This resulted in new estimated correlations for the CATS vs. records based PCL-Rs of 0.95 (SOR), 0.97 (BH) and 0.95 (KR) (Table 1). Whilst these results look encouraging referral to the original correlations of the raw CATS score with the PCL-R score illustrates that at least for the sample calculated by KR this cannot be the entire explanation. It was therefore decided that our sample for the remaining analyses would include only those of the first rater (SOR) giving a sample size of 100.

3.3 What is the nature of the difference in VRAG scores calculated using the CATS and the PCL-R?

3.3.1 In Interview based PCL-Rs?

3.3.1.1 What is the difference in VRAG scaled scores calculated using the CATS rather than the interview based PCL-R?

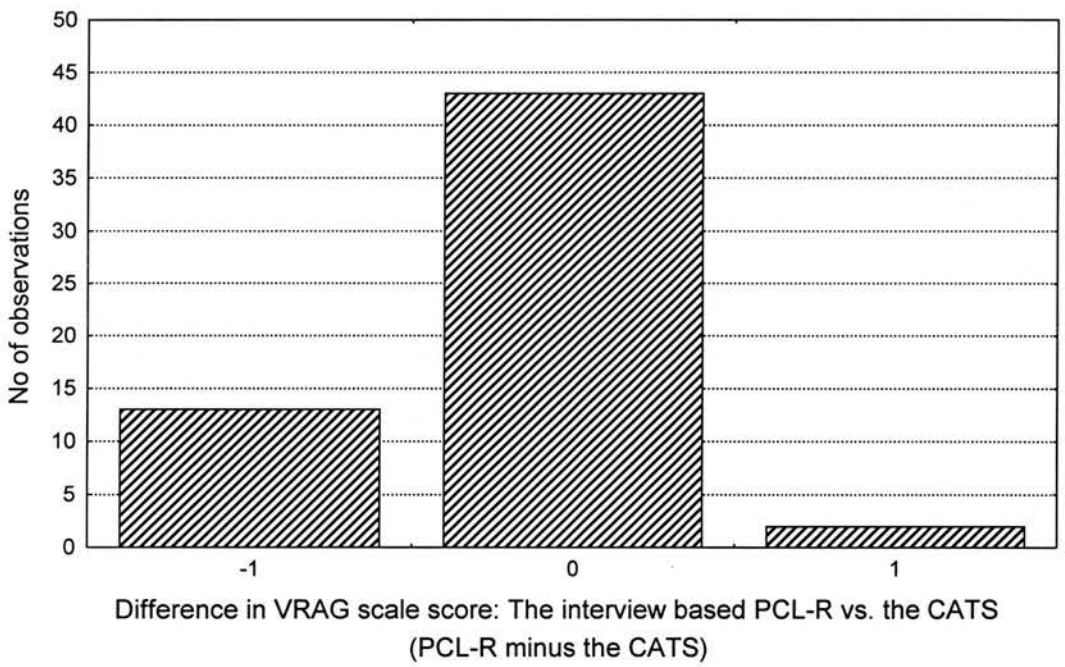
To clarify the nature of the differences between VRAG scores calculated using the interview based PCL-R and the CATS, we noted the frequency of each value of score difference. Calculating the VRAG using the CATS rather than interview based PCL-R score resulted in the same scaled score in 74% of cases with 22% falling one scale score below and 3% one scaled score above the VRAG score as calculated using the PCL-R (Table 9; Figure 5).

Table 9: The distribution of score differences of VRAGs scored using interview based PCL-Rs in comparison to the CATS.

	Error margin: CATS vs. PCL-R	Count	Percent
-1	VRAG calculated using the CATS was one scaled score below that using the PCL-R	13	22.4%
0	No difference between VRAGs calculated using the CATS and the PCL-R	43	74.1%
1	VRAG calculated using the CATS was one scaled score above that using the PCL-R	2	3.4%

Note: Percentages are rounded and therefore may not total 100%

Figure 5: Histogram illustrating the difference in scaled VRAG score when it was calculated using the CATS rather than interview based PCL-R score.



3.3.1.2 Is the validity of the CATS as a replacement for the interview based PCL-R maintained across VRAG score ranges?

To examine if the finding was robust across the range of scaled scores, scores were grouped into two ranges. The correlation between the VRAG as calculated using the CATS in comparison to interview based PCL-Rs for VRAG scaled score range 1-5 was 0.93 (P=<0.05; n=33) and for scores 6-9 was 0.7 (P=<0.05; n=25). Descriptive statistics on each score range were also calculated (Table 10).

Table 10: The reliability of the VRAG calculated using the CATS rather than interview based PCL-R score in two different score ranges.

VRAG Scaled Score	n	Correlation	Dis-attenuated Correlation	Frequency		Mean	Standard Deviation
1-5	33	0.93		-1	4	4.12	0.93
				0	27		
				1	2		
6-9	25	0.74	0.82	-1	9	6.64	0.7
				0	16		
				1	0		

These results suggested that the accuracy of the CATS at replacing the PCL-R in the VRAG may alter across VRAG scaled score ranges as the comparative correlations differed significantly by 0.19 (P=<0.0117; 2 tailed). However, before accepting this as a significant finding, it is noteworthy that due to the smaller sample size in the VRAG scaled score range of 6-9 the standard deviation is severely restricted and too great a homogeneity in a sample can lead to a lowered correlation. Therefore the correlation of the restricted sample was corrected for restricted variance using the 1-5 range variance as the 'population' value yielding a dis-attenuated correlation of 0.82. The corrected correlation, as it is an estimate, with no known distribution, cannot now be tested for significance in comparison with the correlation found in the VRAG scaled score range of 1-5. If, for interest, we anyway

compute the test of significance between the correlations in the two VRAG scaled score ranges (0.93 vs. 0.82), the difference is no longer significant (P=<0.08; two tailed).

3.3.1.3 The Accuracy of predicted VRAG scaled scores using the CATS in comparison to the interview based PCL-R

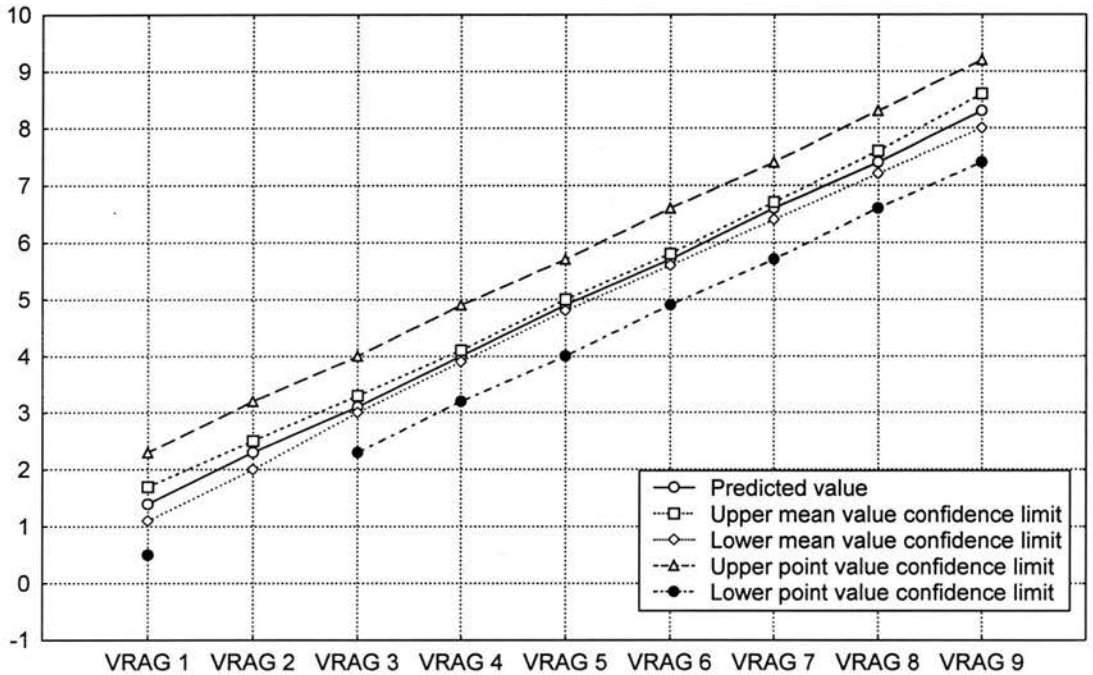
Univariate regression analyses were conducted to allow the prediction of each VRAG scaled score using the CATS with both the point and mean value confidence intervals (Table 11; Figure 6). Point confidence intervals are those applicable to a specific predicted value for an individual case whilst mean confidence intervals refer to those computed around the expected mean score for all possible cases at a particular value of the independent variable.

Table 11: Predicted Scaled VRAG scores using the CATS with point and mean value confidence intervals for interview based PCL-Rs.

VRAG Scaled Score	VRAG Predicted Scaled Score using the CATS	95% Confidence Intervals	
		Point CI	Mean Value CI
1	1.43	0.5 - 2.3	1.1 - 1.7
2	2.29	1.4 - 3.2	2.0 - 2.5
3	3.15	2.3 - 4.0	3.0 - 3.3
4	4.01	3.2 - 4.9	3.9 - 4.1
5	4.87	4.0 - 5.7	4.8 - 5.0
6	5.73	4.9 - 6.6	5.6 - 5.8
7	6.59	5.7 - 7.4	6.4 - 6.7
8	7.45	6.6 - 8.3	7.2 - 7.6
9	8.31	7.4 - 9.2	8.0 - 8.6

The difference between predicted VRAG scaled scores using the CATS and that found using the interview based PCL-R ranges from 0.01 to 0.69 with mean value 95% confidence intervals no greater than 1.0 scaled score remote from the predicted VRAG scaled score and point 95% confidence intervals no greater than 1.7 remote.

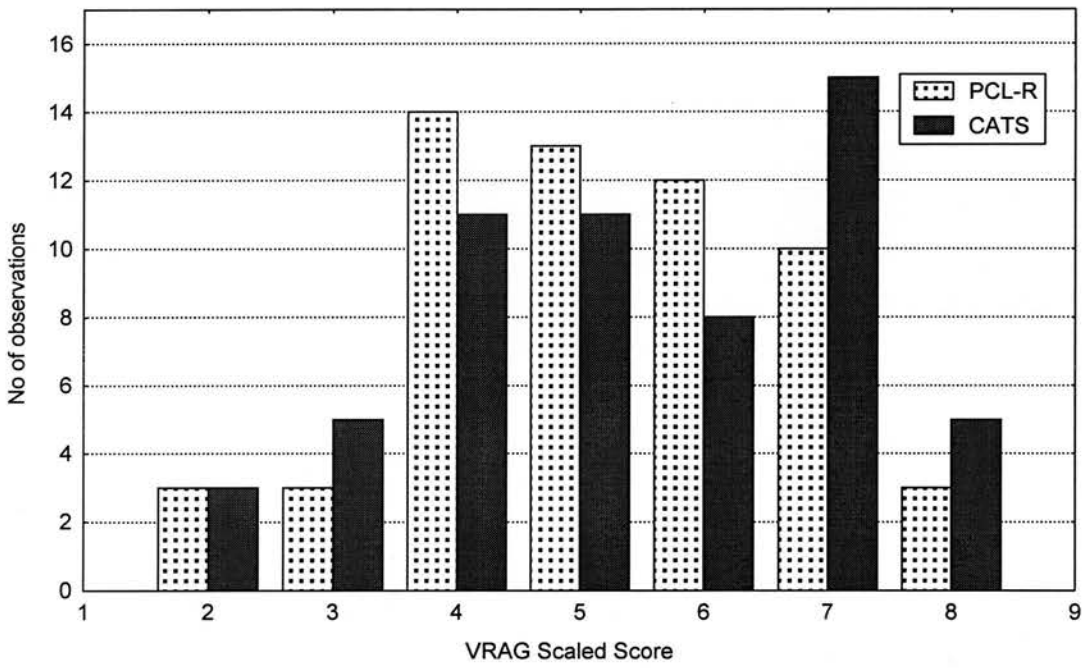
Figure 6: Line plot illustrating predicted interview based PCL-R VRAG scores using the CATS with point and mean value confidence intervals.



3.3.1.4 The Comparative Distribution of total VRAG scores calculated using the CATS in comparison to the interview based PCL-R.

The comparative distribution of total VRAG scores calculated using the CATS and interview based PCL-R scores was plotted in histogram form (Figure 7). The distribution of scores was not significantly different (Chi Sq = 7.451, N-1 categories, ranged from 2-8, df=6, P<=0.28)

Figure 7: The distribution of total VRAG scores calculated using the CATS and interview based PCL-R scores.



3.3.2 In records based PCL-Rs?

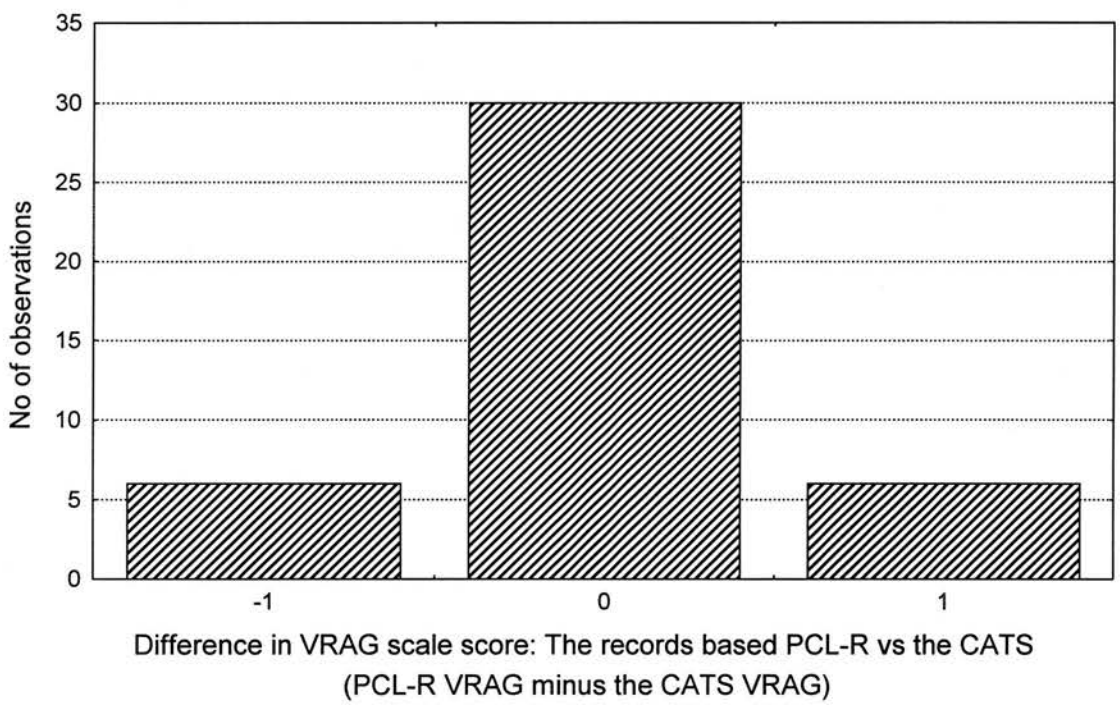
3.3.2.1 What is the difference in VRAG scaled scores calculated using the CATS rather than the records based PCL-R?

To examine the difference between VRAG scores calculated using records based PCL-Rs and the CATS we noted the frequency of each value of score difference. VRAGs computed using the CATS rather than a records based PCL-R score resulted in the same scaled score in 71% of cases with 14% falling both one scale score below and above the VRAG score as calculated using the PCL-R (Table 12; Figure 8).

Table 12: The distribution of score differences of VRAGs scored using records based PCL-Rs in comparison to the CATS.

	Error margin: CATS vs. PCL-R	Count	Percent
-1	VRAG calculated using the CATS was one scaled score below that using the PCL-R	6	14.3%
0	No difference between VRAGs calculated using the CATS and the PCL-R	30	71.4%
1	VRAG calculated using the CATS was one scaled score above that using the PCL-R	6	14.3%

Figure 8: Histogram illustrating the difference in scaled VRAG score when it was calculated using the CATS rather than records based PCL-R score.



3.3.2.2 Is the validity of the CATS as a replacement for the records based PCL-R maintained across VRAG score ranges?

To investigate whether this finding was uniform across the range of scaled scores, scores were again grouped into two ranges. The correlation between the VRAG as calculated using the CATS in comparison to records based PCL-Rs for VRAG scaled score range 1-5 was 0.86 (P=<0.05; n=26) and for scores 6-9 was 0.58 (P=<0.05; n=16). Descriptive statistics on each score range were also calculated (Table 13). As was the case in our analysis of the interview based PCL-Rs the correlations in the two score ranges appeared to differ substantially although in this case the difference failed to reach statistical significance in initial analyses (P=<0.069; two tailed). The higher score range again suffered from a restricted range as illustrated in its low standard deviation and was again dis-attenuated to compensate for homogeneity of variance giving an adjusted correlation of 0.80 which, if we extend the analogy of an adjusted correlation, did not differ from that of the unrestricted sample (0.86 vs. 0.80; P=<0.057; two tailed).

Table 13: The reliability of the VRAG calculated using the CATS rather than records based PCL-R score for two alternative score ranges.

VRAG Scaled Score	n	Correlation	Dis-attenuated Correlation	Frequency		Mean	Standard Deviation
1-5	26	0.86		-1	4	4.19	0.981
				0	19		
				1	3		
6-9	16	0.58	0.80	-1	2	6.63	0.619
				0	11		
				1	3		

3.3.2.3 The Accuracy of predicted VRAG scaled scores using the CATS in comparison to the PCL-R

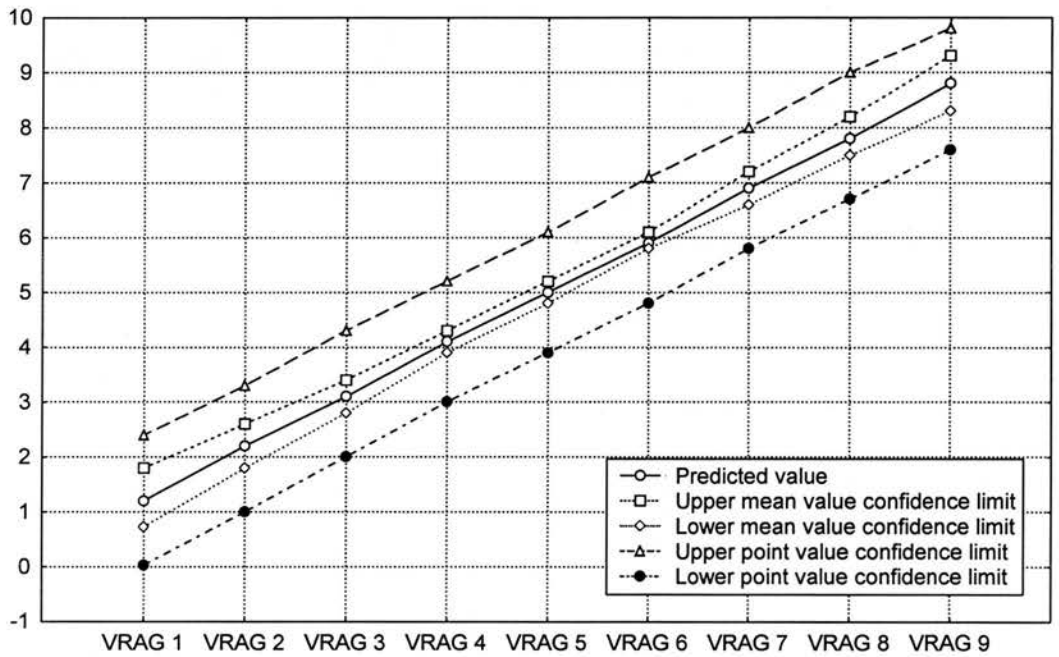
As for interview based PCL-Rs, univariate regression analyses were conducted to allow the prediction of each VRAG scaled score using the CATS with point and mean value confidence intervals (Table 14; Figure 9).

Table 14: Predicted Scaled VRAG scores using the CATS with point and mean value confidence intervals in records based PCL-Rs.

VRAG Scaled Score	VRAG Predicted Scaled Score using the CATS	95% Confidence Intervals	
		Point CI	Mean Value CI
1	1.24	0.03 - 2.4	0.7 - 1.8
2	2.18	1.0 - 3.3	1.8 - 2.6
3	3.12	2.0 - 4.3	2.8 - 3.4
4	4.06	3.0 - 5.2	3.9 - 4.3
5	5.01	3.9 - 6.1	4.8 - 5.2
6	5.95	4.8 - 7.1	5.6 - 6.1
7	6.89	5.8 - 8.0	6.6 - 7.2
8	7.83	6.7 - 9.0	7.5 - 8.2
9	8.88	7.6 - 9.8	8.3 - 9.3

The difference in predicted scaled VRAG scores using the CATS in comparison to those calculated using the records based PCL-R ranged from 0.01 to 0.17, with 95% point confidence intervals of between 0.2 and 0.8 remote from the predicted value and 95% mean value confidence limits of between 1.1 and 1.7 remote from the prediction.

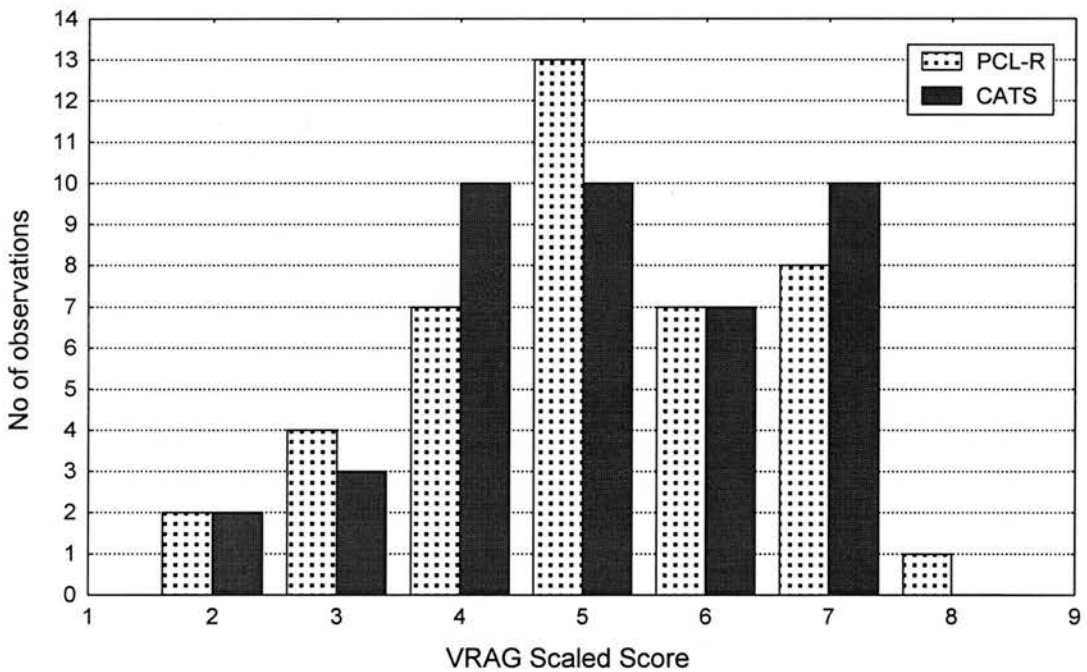
Figure 9: Line plot illustrating predicted records based PCL-R VRAG scores using the CATS with point and mean value 95% confidence intervals.



3.3.2.1 The Comparative Distribution of total VRAG scores calculated using the CATS in comparison to the records based PCL-R.

The comparative distribution of total VRAG scores calculated using the CATS and records based PCL-R scores was plotted in histogram form (Figure 10). The distribution of scores was not significantly different (Chi Sq = 2.339, N-1 categories, ranged from 2-7, df=5, P<=0.80 NS).

Figure 10: The distribution of total VRAG scores calculated using the CATS and records based PCL-R scores.



3.3.3 For the total sample?

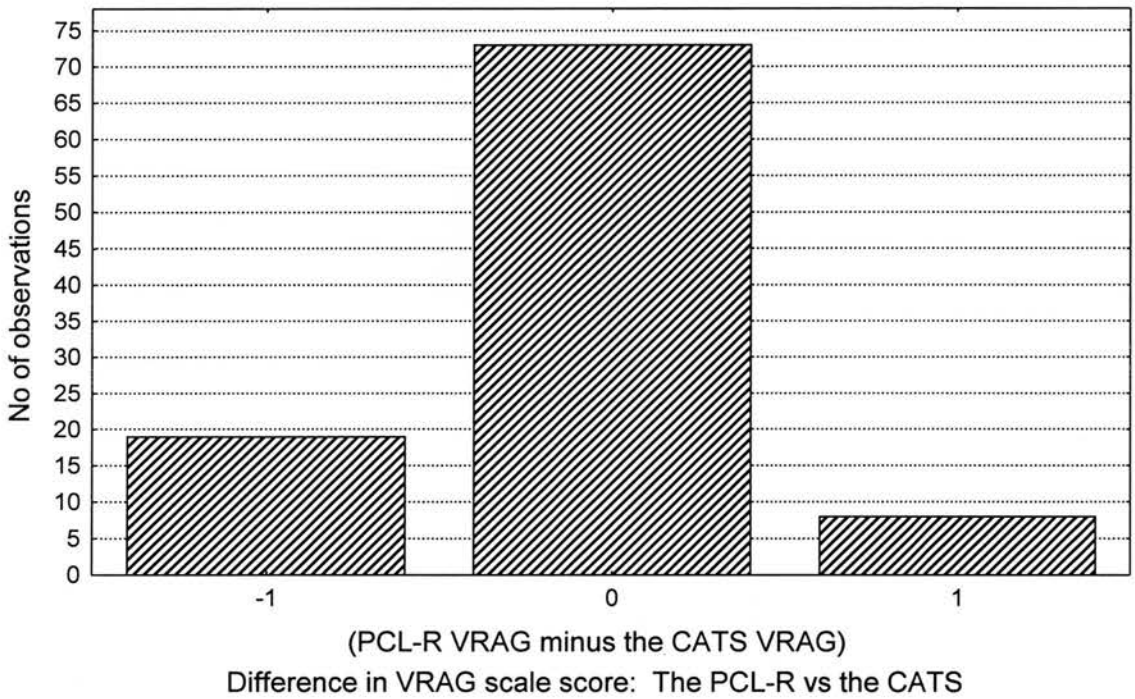
3.3.3.1 What is the difference in VRAG scaled scores calculated using the CATS rather than the combined sample PCL-R?

The distribution of the difference in scores calculated using the CATS in comparison to the PCL-R for the entire sample was then calculated (Table 15) and plotted on a histogram (Figure 11). The CATS accurately replicated the VRAG score achieved using the PCL-R in 73% of cases, estimating one scaled score above in 19% of cases and one scaled score below in 8% of cases.

Table 15: The distribution of score differences of VRAGs scored using the PCL-R in comparison to the CATS.

	Error margin: CATS vs. PCL-R	Count	Percent
-1	VRAG calculated using the CATS was one scaled score below that using the PCL-R	19	19%
0	No difference between VRAGs calculated using the CATS and the PCL-R	73	73%
1	VRAG calculated using the CATS was one scaled score above that using the PCL-R	8	8%

Figure 11: Histogram illustrating the difference in scaled VRAG score when it was calculated using the CATS rather than the PCL-R score.



3.3.3.2 Is the validity of the CATS as a replacement for the combined sample PCL-R maintained across VRAG score ranges?

The reliability of the CATS as a replacement for the PCL-R in the VRAG across two alternative score ranges was then calculated (Table 16). As found previously when examining two score ranges the standard deviations indicated restricted variance in the higher range and again this was adjusted giving correlations and percentage errors of 0.90, 22% (range 1-5) and 0.77, 34% (adjusted; range 6-9). If we extend the analogy of the estimated correlation to examine the difference between the two correlations they are found to be significant (P=<0.03; n=100) although the reliability of doing so is limited. Therefore it appears that in the higher score ranges the CATS is beginning to over estimate the VRAG.

Table 16: The reliability of the VRAG calculated using the CATS rather than the combined sample of PCL-R scores for two alternative score ranges.

VRAG Scaled Score	n	Correlation	Dis-attenuated Correlation	Frequency		Mean	Standard Deviation
1-5	59	0.90		-1	8	4.15	0.943
				0	46		
				1	5		
6-9	41	0.65	0.77	-1	11	6.63	0.662
				0	27		
				1	3		

3.3.3.3 The Accuracy of predicted VRAG scaled scores using the CATS in comparison to the combined sample of PCL-R scores

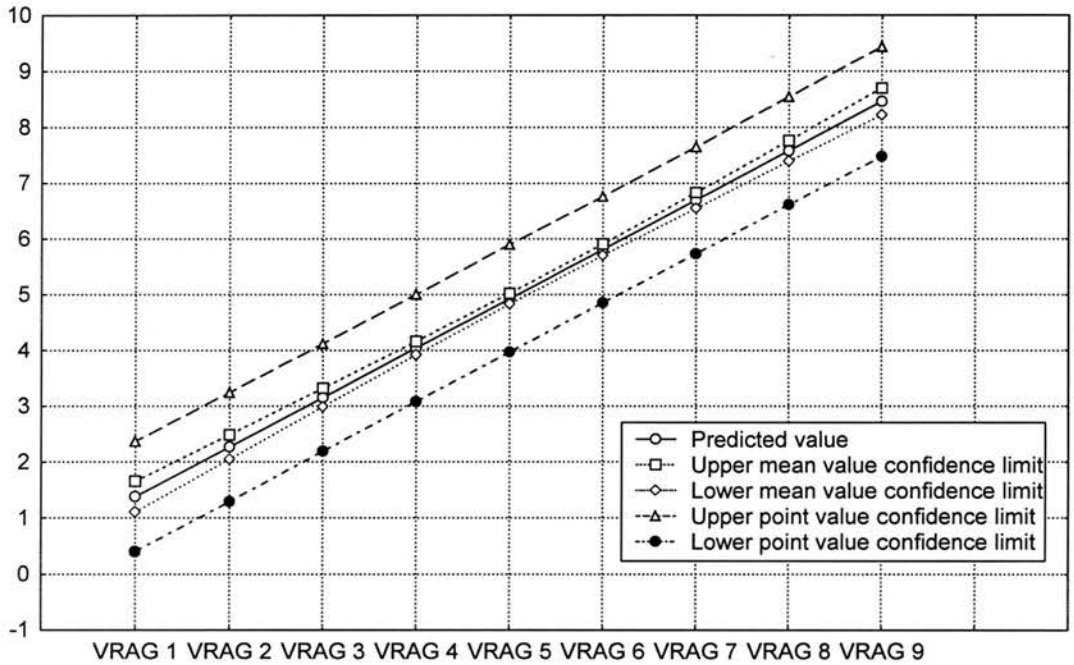
As previously conducted on records and interview based samples, univariate regression analyses were conducted to allow the prediction of each VRAG scaled score using the CATS with both the point and mean value confidence intervals for the combined sample (Table 17; Figure 12).

Table 17: Predicted Scaled VRAG scores using the CATS with point and mean value confidence intervals for the combined sample.

VRAG Scaled Score	VRAG Predicted Scaled Score using the CATS	95% Confidence Intervals	
		Point CI	Mean Value CI
1	1.38	0.4 – 2.4	1.1 – 1.7
2	2.27	1.3 – 3.2	2.1 – 2.5
3	3.15	2.2 – 4.1	3.0 – 3.3
4	4.04	3.1 – 5.0	3.9 - 4.2
5	4.92	4.0 - 5.9	4.8 - 5.0
6	5.81	4.9 – 6.7	5.7 – 5.9
7	6.69	5.7 – 7.7	6.6 – 6.8
8	7.58	6.6 – 8.5	7.4 – 7.7
9	8.46	7.5 – 9.4	8.2 – 8.7

The difference between predicted VRAG scaled scores using the CATS and that found using the PCL-R for the combined samples ranges from 0.04 to 0.54 with mean value 95% confidence intervals no greater than 1.3 scaled scores remote from the predicted VRAG scaled score and point 95% confidence intervals no greater than 1.6 remote.

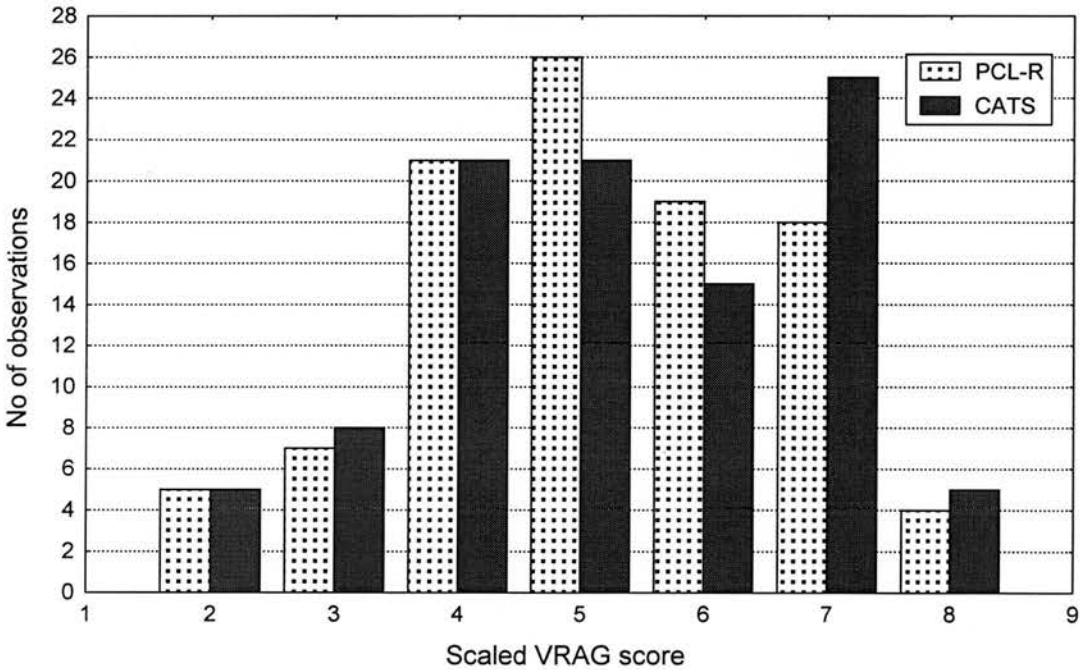
Figure 12: Line plot illustrating predicted interview based PCL-R VRAG scores using the CATS with point and mean value confidence intervals.



3.3.3.4 The Comparative Distribution of total VRAG scores calculated using the CATS in comparison to the combined sample PCL-R.

The distribution of total VRAG scores calculated using the CATS in comparison to those calculated using records based PCL-R scores was plotted in histogram form (Figure 13). Their distribution did not differ significantly (Chi Sq = 4.919, N-1 categories, ranged from 2-8, df=6, P<=0.55 NS).

Figure 13: The distribution of total VRAG scores calculated using the CATS and combined PCL-R scores.



3.4 How does the CATS replicate the PCL-R scores, in what way are their composite items related?

To examine how each CATS item related to each other, the scale's internal structure, the correlation of each variable with all others, was calculated (Table 18).

Table 18: Correlation of each CATS variable and totals against all others.

CATS Variable	1	2	3	4	5	6	7	8	Total
1	1	**0.28**	**0.44**	**0.50**	**0.64**	**0.55**	0.14	**0.42**	**0.79**
2	**0.28**	1	0.13	**0.22**	**0.32**	**0.26**	**0.25**	0.06	**0.49**
3	**0.44**	0.13	1	0.15	**0.30**	0.19	0.09	0.17	**0.48**
4	**0.50**	**0.22**	0.15	1	**0.52**	**0.60**	0.07	**0.46**	**0.70**
5	**0.64**	**0.32**	**0.30**	**0.52**	1	**0.56**	0.19	**0.37**	**0.78**
6	**0.55**	**0.26**	0.19	**0.60**	**0.56**	1	0.18	**0.36**	**0.74**
7	0.14	**0.25**	0.09	0.07	0.19	0.18	1	0.18	**0.42**
8	**0.42**	0.06	0.17	**0.46**	**0.37**	**0.36**	0.18	1	**0.61**
Total	**0.79**	**0.49**	**0.48**	**0.70**	**0.78**	**0.74**	**0.42**	**0.61**	1

Significant correlations are in bold.

An understanding of the comparative distribution of each sample's scaled score was gained by adjusting the records based scores (n=42) in proportion to the percentage difference in sample size in comparison to the interview based sample (n=58). Due to the small numbers in some groups scaled scores were grouped in the ranges 1-3, 4-6 and 7-9. The distribution of expected versus observed frequencies was examined using the Chi Square statistic and did not reach statistical significance (P=<0.069) although indicating there was a borderline difference in the distribution of the score ranges.

The relationship between each CATS variable and each PCL-R variable and its factors was examined by again constructing a correlational table comparing all variables (Table 12). The comparative distribution of total CATS and PCL-R scores was plotted on a histogram (Figure 11).

Table 19: Correlation of each CATS variable with each PCL-R variable for the sample as a whole.

Variable	CATS 1	CATS 2	CATS 3	CATS 4	CATS 5	CATS 6	CATS 7	CATS 8	CATS Total
PCL-R 1	0.16	0.12	0.09	0.08	-0.02	-0.00	-0.05	0.17	0.11
PCL-R 2	0.03	0.03	-0.01	-0.02	0.06	-0.00	**-0.27**	0.01	-0.04
PCL-R 3	**0.34**	**0.26**	0.15	**0.35**	**0.36**	**0.22**	-0.01	**0.21**	**0.38**
PCL-R 4	0.19	0.18	-0.03	0.09	0.00	0.11	0.05	0.04	0.13
PCL-R 5	0.19	0.10	-0.05	0.18	0.12	**0.24**	0.09	**0.23**	**0.22**
PCL-R 6	0.11	0.05	-0.11	0.13	0.11	0.11	**-0.22**	-0.11	0.01
PCL-R 7	0.12	-0.00	-0.09	0.14	0.06	0.03	-0.20-	0.11	-0.01
PCL-R 8	0.08	0.00	-0.10	0.07	0.08	-0.03	**-0.23**	0.01	-0.02
PCL-R 9	0.08	0.11	-0.04	0.16	0.12	0.16	-0.12	0.05	0.10
PCL-R 10	0.15	0.13	-0.07	0.14	**0.24**	**0.22**	0.03	0.18	**0.21**
PCL-R 11	-0.05	0.07	-0.08	0.11	0.06	0.12	0.10	0.12	0.09
PCL-R 12	**0.42**	0.19	0.20	**0.30**	**0.44**	**0.36**	0.12	**0.34**	**0.48**
PCL-R 13	0.09	-0.06	0.000	0.11	0.00	0.09	-0.11	0.03	0.03
PCL-R 14	0.07	0.17	0.02	**0.21**	0.18	0.16	**0.23**	0.19	**0.25**
PCL-R 15	0.06	0.09	-0.17	**0.25**	0.12	0.01	0.00	0.16	0.11
PCL-R 16	0.09	-0.08	-0.13	0.04	0.02	-0.09	-0.05	0.02	-0.04
PCL-R 17	-0.13	0.17	0.02	-0.09	0.02	-0.07	0.15	0.11	0.04
PCL-R 18	**0.32**	**0.29**	0.27	**0.40**	**0.51**	**0.42**	0.10	**0.31**	**0.53**
PCL-R 19	**0.44**	**0.29**	0.32	**0.36**	**0.47**	**0.31**	-0.03	**0.28**	**0.49**
PCL-R 20	**0.36**	**0.35**	0.25	**0.29**	**0.49**	**0.37**	**0.24**	0.20	**0.51**
Total	**0.34**	**0.26**	0.06	**0.36**	**0.38**	**0.31**	-0.01	**0.26**	**0.39**

Note: N=94 as the result of missing data on the PCL-R.
Significant correlations are in bold.

To further clarify how individual CATS and PCL-R items were related we conducted a factor analysis of them as a joint item set. We began by conducting three tests of factor extraction quantity using both Principle Component (PCA) and Image Component (IFA) Analyses to determine the number of factors contained in the data set. These were the

Velicer Minimum Average Partial (MAP) Test, Armor's Theta and Barrett and Kline's Autoscree from which we concluded that there should be two factors (Table 20).

Table 20: The results of the three extraction quantity tests to determine the number of factors in the sample (n=100).

Method	Principle Component Analyses	Image Component Analyses
MAP Test	2	2
Theta Reliability	3*	2
Autoscree	5	2

Borderline (the 3rd theta was 0.53).

The amount of variance identified as common variance (item image variance) within the correlation matrix was 57%. Taking into account the results of the factor extraction quantity tests (Table 20) we analysed the data using a 2 factor minimum residual analysis (MINRES) and a direct oblimin rotation. For the rotation delta was swept in 0.5 increments from -40.0 to 0.5. Although the theoretical maximum of delta is 1.0 we did not exceed 0.5 above which factor solutions tend to collapse because of extreme obliquity. The maximal simple structure was found at delta = 0.5, with an overall hyperplane count of 20 (as defined by loadings within $\mp$ 0.1 where the maximum would be 56, 28 variables multiplied by 2 dimensions). The factor correlation was 0.134 which is only mildly oblique suggesting the factors are semi-independent. The two factors and the individual loadings of variables on each are listed in Table 21.

Table 21: The combined factor pattern of the combined CATS and PCL-R variables with the loadings representing the contribution of the factor to each variable, independent of the correlation between factors.

Variable		Factor 1	Factor 2	PCL-R Factor
CATS 1	Elementary school maladjustment	**0.688**	0.042	
CATS 2	Teenage alcohol problem	**0.392**	0.031	
CATS 3	Childhood aggression rating	**0.421**	-0.196	
CATS 4	Conduct Disorder symptoms	**0.619**	0.092	
CATS 5	School suspension / expulsion	**0.774**	-0.004	
CATS 6	Arrested under 16 years	**0.653**	0.003	
CATS 7	Parental alcoholism	0.257	-0.170	
CATS 8	Lived with both parents to age 16?	**0.508**	-0.001	
PCL-R 1	Glibness / Superficial charm	-0.28	**0.470**	1
PCL-R 2	Grandiose sense of self worth	-0.004	**0.334**	1
PCL-R 3	Need for stimulation / Proness to boredom	**0.456**	**0.362**	2
PCL-R 4	Pathological lying	0.088	**0.488**	1
PCL-R 5	Conning / Manipulative	0.98	**0.533**	1
PCL-R 6	Lack of remorse or guilt	-0.098	**0.762**	1
PCL-R 7	Shallow affect	-0.062	**0.558**	1
PCL-R 8	Callous / Lack of empathy	-0.134	**0.831**	1
PCL-R 9	Parasitic Lifestyle	0.036	**0.302**	2
PCL-R 10	Poor behavioural controls	0.197	**0.558**	2
PCL-R 11	Promiscuous sexual behaviour	0.167	0.010	
PCL-R 12	Early behavioural problems	**0.628**	0.143	2
PCL-R 13	Lack of realistic, long term goals	-0.041	**0.459**	2
PCL-R 14	Impulsivity	0.239	**0.372**	2
PCL-R 15	Irresponsibility	0.55	**0.749**	2
PCL-R 16	Failure to accept responsibility for own actions	-0.144	**0.663**	1
PCL-R 17	Many short-term marital relationships	0.059	-0.099	
PCL-R 18	Juvenile Delinquency	**0.728**	0.054	2
PCL-R 19	Revocation of conditional release	**0.625**	-0.082	2
PCL-R 20	Criminal versatility	**0.617**	-0.121	
Hyperplane count		10	10	
Variance		4.662	4.471	

Loadings above 0.3 are in bold.

The factor analyses suggest that the CATS is a well defined scale, loading strongly onto one factor, which taps into the PCL-Rs measures of childhood behaviour and criminality, the two areas that appear to best describe the nature of Factor 1. Factor two is similar to the PCL-R's traditional first factor reflecting personality characteristics. All PCL-R items normally

contained within its first factor reflecting 'selfish, callous, and remorseless use of others', fall into the combined CATS PCL-R Factor 2 reflecting personality characteristics. Whilst the majority of items traditionally in the second PCL-R factor reflecting 'Chronically unstable, anti-social, and socially deviant lifestyles' are also in the combined model's factor 2 the strength of the new Factor 1 dominated by the CATS is such that it has 'dragged' similar PCL-R items across to join it.

Note:
For information the frequency of each response for each question of the VRAG (Appendix 4), CATS (Appendix 5) and PCL-R (Appendix 6) for interview based, records based and the combined sample PCL-R are provided in the appendices.

4 Discussion

4.1 Is the CATS a valid alternative to the PCL-R in calculating the VRAG?

The purpose of this study was, in the light of increasing, legislation based, pressure on clinicians to accurately predict risk, to examine the validity of a quicker, more effective means of calculating the VRAG, presently a widely used method of actuarial risk prediction.

The current findings suggest that the CATS is a valid alternative to the PCL-R score in calculating the VRAG with little loss of accuracy in a UK forensic inpatient population. Its replacement of the PCL-R would have important benefits in negating the need for a specially trained PCL-R rater, or for a structured interview, giving time and cost benefits. The VRAG scaled score using the CATS was identical to that using the PCL-R in 73% of cases achieving a correlation of 0.95. In no cases did the VRAG scaled score using the CATS differ by more than one scaled score to that produced using the PCL-R indicating a minimal margin of error.

The examination of the accuracy of the CATS in predicting VRAG scores calculated conventionally with the PCL-R illustrated different levels of accuracy in different conditions. For the primary rater in comparison to interview based PCL-Rs the correlation was highest at 0.96 but dropped marginally to 0.93 in comparison to records based PCL-Rs. The reasons for this may be twofold. Perhaps PCL-Rs calculated by records only, are inherently less reliable than those inclusive of an interview. If this is so, it is of note that the CATS better approximates the more accurate interview based measure. The PCL-R manual states that ratings can be made using case records only, providing that they are of sufficient quality, but does not report a reduction in reliability. This brings us to our alternative explanation, the

reduction in accuracy may be a product of the subjects in each group and the quality of information available on them. Subjects in the interview based PCL-R group were currently resident within the State Hospital whereas those in the records based group were persons discharged from the hospital between 1985 and 1987.

It was the experience of the primary rater (SOR) that there was less identifying information from which to score the VRAG in the latter group. Files for subjects in the latter historical group lacked the structured report formats of the more recently referred patients. Difficulties arose where patients admitted more than a decade previously lacked the detailed accounts of childhood and adolescent history so integral to modern assessment. If this difficulty was also experienced when the PCL-R was being calculated for these subjects this may have further contributed to reduced reliability and hence the reduced correlation between VRAGs calculated using these PCL-R scores versus the CATS in comparison to interview based questionnaires. This point also illustrates a potential problem with both the CATS and the VRAG, their dependence on accurate historical data.

4.2 The influence of raters

The ability of the CATS to replicate the VRAG calculated using the PCL-R score calculated on records also differed according to rater dropping from 0.93 (SOR) to 0.88 (BH) and 0.81 (KR). As noted during our analyses, this pattern of results may simply reflect restricted variance as a result of small sample sizes and indeed the dis-attenuated correlations, altering these correlations to 0.95 (SOR & KR) and 0.97 (BH), indicated that this may be the case. However for a dis-attenuated correlation to be accurate a number of assumptions must be made regarding the sample including linearity and bivariate normality, the lack of which may render the estimate inaccurate. Reference to the correlations between raw CATS score

and raw PCL-R scores where rater KR's was -0.3 suggests that homogeneity of variance may not be entirely responsible for the difference in correlations between raters. If this is the case an alternative explanation may return us to the different nature of the subjects in the records based and interview based PCL-R groups as these correlations all refer to subjects whose records were of the older type. All three raters reported difficulty locating relevant information and that some simply was not contained in the files necessitating rater 'judgement calls', their best guess (see 4.2.1). It may be that the difference in the CATS' accuracy reflects the clinical experience of the raters as the reducing correlations mirror the number of years of clinical experience of each rater, 7 years (SOR), 3 years (BH) and 1 year (KR). Thus there may have been an interaction between the poor quality of the files and the need for 'judgement calls' and clinical experience. It could of course be argued that if this is the case it negates the very benefits the CATS was thought to hold over the PCL-R, that of a time and cost saving as a clinically experience trained rater was not required. This can be countered on two fronts. Firstly, this finding requires to be replicated on a larger sample to allow the separation of any effects of restricted variance due to small sample sizes. Secondly, it needs to be replicated on more modern, accurate hospital files to identify if the reduced need for 'judgement calls' would remove the effect.

4.2.1 Objectivity in scoring the CATS and the VRAG

Ideally the clinical experience of the rater should not be an issue for an actuarial instrument where variables should be precisely defined and the answer a matter of fact rather than opinion.

In the present study it was apparent that this was not so (see 4.2.). A detailed examination of the VRAG and CATS revealed that there was potential for substantial response variation.

For example:

The first question regarding if the subject lived with both biological parents until the age of 16 raised a number of questions.

- Did separation during a mother's confinement for a subsequent child count?
- If being sent to an approved school counted as separation did going to boarding school?
- How long a time period constituted separation?

Thus there were many questions regarding the definition of a superficially straightforward variable. Much more difficult were questions regarding elementary school maladjustment or childhood aggression, the types of terms that are by nature difficult to categorise, especially retrospectively from case note data. In an attempt to make decisions uniform, as each query of this nature arose a decision was taken after consultation with the original authors and recorded for future reference (Appendix 3). I believe that these additional notes regarding coding may provide a valuable tool for those wishing to use the VRAG in the future. 'Judgement calls' were therefore limited to decisions of whether the absence of information in a file was a true indication of an issue not being a problem. For example, if parental alcoholism is not mentioned as being either present or absent, do we assume the latter? The VRAG scoring guide suggests this approach is taken but, most importantly, has the caveat that the file should be of a certain standard before one can assume this is the case. It is notable that prior to the VRAG's development considerable effort was spent re-structuring and ensuring the quality of the hospital files from which it was subsequently scored. Such an initiative had not taken place at the State Hospital resulting in the difficult 'judgement calls' previously noted. Rice suggests that should there be insufficient information to confidently score an item that it should be omitted and no score given (Rice, personal communication).

We did not adopt this approach, instead choosing a 'forced choice' option, as reference to the score weighting system on the VRAG indicates that a score of zero for an absent item, rather than contributing nothing, actually biases the instrument toward a higher score, indicating greater likelihood of recidivism, as many indicators of reduced recidivism on the VRAG are weighted with a negative score.

4.3 What is the nature of the errors in CATS calculated VRAGs?

A further difference in the accuracy of VRAGs calculated using the CATS in relation to interview or records based PCL-Rs was the distribution of error scores. Whereas for the records based PCL-Rs the CATS VRAG errors were evenly distributed at 14% estimating one scaled score below and above, those for interview based PCL-Rs were 22% one scale score below and 3% one scaled score above. As difference scores were calculated as the PCL-R VRAG minus the CATS VRAG, this suggests that for interview based PCL-R VRAGs the CATS VRAG errors seem to be skewed towards an overestimation of likelihood to recidivate. That is, in 22% of cases the VRAG calculated using the CATS was estimated at one scaled score point higher than using the interview based PCL-R. Thus if using the CATS a trend may exist toward over-estimating someone's potential to recidivate and a higher rate of false positives.

It may be that raters dependent on historical records only are forced to record as present evidence of a personality PCL-R item, that is mentioned once in the notes, that a clinician may dismiss if unable to elicit at interview. This again makes the assumption that records based PCL-Rs are marginally less accurate than their interview based counter-parts. Thus,

although we could simply conclude that the CATS is a more valid predictor of records based PCL-Rs, it is more clinically ethical to aspire to the more accurate measure and we should therefore endeavour to replicate the findings of interview based PCL-Rs. If the relationship appears to be uniform and replicable in a larger sample a re-working of the weighting system of the CATS to more closely approximate results gained using the interview based PCL-R may be desirable. Yet it should be borne in mind that at no time were VRAG scaled scores using the CATS more than one scaled score different to those gained using the PCL-R.

Of similar concern is if the relationship of the CATS to the PCL-R within the VRAG was not uniform across the scale, as this again would require corrective weighting. Initial analyses of grouped VRAG scaled scores into ranges of 1-5 and 6-9 appeared to suggest that the CATS was a less valid replacement of the PCL-R in the VRAG's higher score ranges for interview based, records based and combined PCL-Rs. However, these data again suffered from increased homogeneity of variance as the result of small samples sizes and once dis-attenuated the difference between CATS vs. PCL-R correlations for the two score ranges was no longer significantly different in the interview based and records based samples. However within these samples the trend toward reduced validity still existed and within the combined sample it remained significant. Whilst in such small samples one cannot be confident of such trends it appears that for the VRAG scaled score higher ranges the CATS may over-estimate the VRAG in comparison to its value when calculated using the PCL-R as noted for interview based PCL-Rs alone. This further replicates the finding of its propensity to overestimate the VRAG scaled score when considering simple frequency counts of error scores across the VRAG range as a whole but suggests that this bias may be skewed toward higher score ranges meaning considerable care would have to be taken in the development of any alternative weighting system. However, it must be stressed that the CATS estimated VRAG did not differ from that calculated using the PCL-R by more than one scaled score for any case.

The question remains why the CATS should replicate the PCL-R less effectively in the higher score ranges. The explanation may be that in the higher score ranges more judgement calls are required as the rating of the absence of a variable, in comparison to its severity, is more easily judged. Alternatively the PCL-R's first factor, identifying personality variables may be tapping into dynamic variables whereas the CATS, being solely historically based, will remain the same independent of when during a person's adulthood it is scored. Thus the PCL-R may be identifying a reduction in risk of violent recidivism.

4.4 In what way are CATS and PCL-R items related?

Examination of the internal structure of the CATS suggested that there were no redundant variables as all correlated significantly with the CATS total score. The variable which correlated most highly with the score total was elementary school maladjustment which had a correlation of more than 0.4 with all variables apart from those referring to teenage and parental alcohol abuse, which appeared to be independent, neither correlating significantly with any other variable. The existence of conduct disorder was highly correlated with suspension or expulsion from school and arrest prior to the age of 16 years.

The literature on individual predictors of violent recidivism has examined few of the variables that appear in the CATS (Bonita et al., 1998) but does add weight to the validity of some. The CATS variables of school maladjustment, school expulsion, childhood aggression, conduct disorder and arrest before the age of 16 years could all be considered sub-categories of a 'pre-adult history of antisocial behaviour' (Gendreau et al., 1996) predictive of general recidivism or 'juvenile delinquency' (Bonita et al., 1998) predictive of violent recidivism. Neither is whether the offender lived with both parents addressed

specifically in the literature although this and parental alcoholism may contribute to 'family structure' predictive of general recidivism (Gendreau et al., 1996) and 'family problems' predictive of violent recidivism (Bonita et al., 1998). The teenage and parental alcoholism variables in the CATS find some parallels in the correlation of alcoholic fathers with later violence (Scott, 1977) and the significant relationship between substance abuse and both general (Gendreau et al., 1996) and violent recidivism (Bonita et al., 1998).

If an eight item actuarial scale can so closely approximate the results of a twenty item clinical scale, for the purposes of VRAG computation, the question must be asked of how the two scales relate to one another. The correlations of variables from each scale against each other and the results of the factor analysis of all twenty-eight items were remarkably consistent. The CATS and PCL-R total scores are significantly positively correlated. Only five PCL-R items correlated significantly with individual CATS items and these five did so consistently being significantly related to at least four CATS variables each. The strongest of these was juvenile delinquency followed by criminal versatility, revocation of a conditional release, early behavioural problems and the need for stimulation / proness to boredom. Factor analysis of the 28 items suggested that they contained two factors, the first of which identified criminal history and childhood maladjustment and the second of which reflected aspects of personality very much mirroring the two known factors of the PCL-R. All eight CATS variables fell within Factor 1, as would be anticipated considering its historical emphasis and reflecting a cohesive structure. Fourteen PCL-R items fell within Factor 2 including all those in the PCL-Rs traditional Factor 1 indicative of 'selfish, callous, and remorseless use of others' (Hare, 1991), again as one would expect as this second combined CATS /PCL-R factor taps into personality items. The remaining five items that contribute to the PCL-Rs second factor reflecting 'chronically unstable, antisocial, and socially deviant lifestyle' (Hare, 1991) (as one question is contained in neither) all therefore load onto factor 1 of the combined factor structure. These are the same five variables that

were significantly correlated with the CATS variables as described earlier which are thus embodied in the factor analysis as one would expect. All relate to childhood maladjustment and criminal history. Thus the CATS is a cohesive scale that primarily reflects the traditional second factor of the PCL-R and replicates its assessment of early maladjustment and criminal propensity.

4.5 The clinical utility of calculating the VRAG using the CATS rather than the PCL-R

The internal structure of the CATS and its relationship with the PCL-R are consistent and its correlation, across the sample, of 0.95 appears impressive but what of it's clinical utility? We have noted that the VRAG calculated using the CATS is never more than one scaled score different to that calculated using the PCL-R but this occurs in 27% of cases and the consequences of this must be considered. Let us take as an example the prediction of a PCL-R VRAG scaled score of four as this is the most common VRAG score both in this and the developmental sample. The predicted score would be 4.04 and the point confidence intervals (if we are considering an individual case) would be from 3.1 to 5.0 indicating, as we would expect, the no more than one scaled score error rate. If we translate this into projected recidivism rates though the implications for clinical utility become apparent. These confidence intervals represent predicted recidivism ranging from 0.12 to 0.36 (for VRAG score 4 = 0.17) over 7 years and 0.24 to 0.48 (for VRAG score 4 = 0.31) over 10 years. The breadth of these ranges of predicted recidivism are such that it's utility to the clinician must be called into question. Thus the VRAG calculated with the CATS rather than the PCL-R introduces considerably more error than their correlation initially indicates. Studies of the VRAG's ability to accurately predict recidivism suggest it is 76% accurate, an impressive

figure in risk assessment but how is this affected when calculated using the CATS? Combining the reduction in variance explained by the CATS score and the VRAG's independent accuracy the altered CATS VRAG accuracy is reduced to 66%, a 10% drop. An ability to predict violent recidivism with 66% accuracy remains an impressive feat when compared to studies of clinical prediction (see 1.6.1.). However, such an accuracy rate suggests that the inclusion of standard errors for use in conjunction with scaled VRAG scores may be useful for clinicians and that there is a cost to using the CATS rather than the PCL-R.

4.6 A proposed method of improving actuarial measurement

The effect of the introduction of actuarial assessment on base rate recidivism has implications for the actuarial instrument's lifespan. By definition, the implementation of an effective actuarial instrument will reduce base rate recidivism as it will suggest the continued detention of those most likely to re-offend. This prevents the 're-norming' of the instrument and only allows for its accuracy in terms of true negatives and false negatives to be assessed. It will not be possible to produce a new actuarial instrument as you have limited those offenders you release. It is therefore necessary to begin follow up immediately to assess if its accuracy amongst those released is as one would predict. Presently, it has been possible to continue to develop actuarial instruments because clinical judgement is less than optimal. The only means by which to improve an actuarial instrument once it has been implemented, other than monitoring its false negative rate, is to attempt to reduce its false positive rate. Yet by definition we cannot determine our false positive rate without manipulating our thresholds of likely recidivism for release, that is releasing those offenders who would not

have been released ordinarily, to monitor their behaviour, something which is ethically dubious. However, it may be possible to devise a second step in the actuarial process where those patients falling above the threshold for release on the actuarial instrument are subject to further assessment to ascertain if they share factors known to be common amongst false positives for this instrument. A proportion of these persons may then become eligible for a form of conditional release thus reducing the instrument's false positive error rate.

The present study proposes a two stage actuarial system.

1) The actuarial instrument provides an initial estimate of recidivism probability for an offender.
2) Those persons not eligible for release are examined with a second actuarial instrument that determines their likelihood of being a false positive allowing the release of a further group.

The instrument could be further refined with ongoing monitoring of its false negative rate.

4.7 Implications of these results

The current study has three primary implications regarding the calculation of the VRAG using the CATS rather than the PCL-R.

The CATS may require re-scaling to allow it more consistently to replicate the PCL-R across VRAG score ranges and to better replicate interview based PCL-Rs. The finding that the CATS VRAG had a lower correlation with the PCL-R VRAG at the upper end of the VRAG scaled score range may simply be a product of our small sample and its restricted range but it

would be desirable to examine this further. Should the CATS genuinely differ in its ability to replicate the PCL-R within the VRAG's higher score ranges then its variable weightings should be altered to allow it to do so more consistently. Such re-scaling to compensate for both this and to allow the more accurate replication of interview PCL-Rs would be a relatively simple task if data were available on a large enough sample.

The CATS scale is substantially less time consuming than the PCL-R and does not require patient access whilst still achieving a high level of accuracy making widespread screening for risk of violent recidivism with an actuarial instrument a more achievable goal. Should further study indicate that our initial finding of a rater experience effect is instead due to small sample sizes or the quality of patient records the CATS would also have the substantial benefit of not requiring a clinically experienced and specifically trained rater as does the PCL-R. This would further reduce the costs involved in its use.

It should be noted that this study has examined the utility of the CATS as a replacement for the PCL-R only for the calculation of a VRAG score. The results of the current study do not suggest, and its developers do not recommend, its replacement of the PCL-R in clinical practice. The CATS is not a measure of psychopathy.

4.8 The relevance of these findings to current clinical practice

Whilst the present study may have offered further evidence of the utility of actuarial methods, its relevance to current clinical practice may be limited a priori by the complete rejection of actuarial methods by all special hospitals in the UK. Broadmoor hospital has

recently prepared a position paper rejecting actuarial methods, and Ashworth hospital is planning to study the actuarial prediction of sexual offender recidivism whilst Rampton is not currently using actuarial measures and has no official policy on their implementation (Hammond, Personal communication; Logan, Personal Communication; Hodge, Personal Communication). None are currently using, or planning to use, actuarial methods for predicting risk, a policy also shared by the Scottish Prison Service (Barrett, Personal Communication). The State Hospital, Carstairs is conducting a replication study of the VRAG working closely with Rice et al. in Canada and is researching the efficacy of actuarial measures as a necessary precursor before considering their implementation. At present actuarial measures are not used routinely in the assessment of risk at the State Hospital.

Thus even though verifiable evidence exists that actuarial methods can be more accurate than current clinical judgement (Quinsey et al., 1998; Meehl, 1954), clinicians are still opposed to their use. The accuracy of risk assessment will be of increasing importance if the recommendations of the Fallon enquiry (Fallon et al., 1999) for a third service and reviewable sentences for those with personality disorders are implemented and is something we must address.

The key to clinicians' apparent reluctance appears to be the creation by actuarial instruments of a written record of the expected probability of recidivism. This has two implications for the clinician who must make decisions regarding release. First, an actuarial score in terms of probabilities does not answer the clinician's dichotomous question regarding release but rather continues to leave them with the choice of what constitutes an acceptable risk, as we cannot detain all persons with a score of above one. Secondly, the record of a numerical probability of recidivism may serve to provide critics unfamiliar with the issues, should recidivism occur, with data with which to challenge the clinician no matter how low that probability figure was, because acceptable risk cannot be objectively defined.

The answer may lie in this very fact. The level of risk of violent recidivism that is acceptable amongst those to be released must be defined by the society that accepts the person back within its ranks. It is a social construction (Bingley, 1997). It is a political decision that should not be given to one person or a group of medical personnel but rather to an independent group that makes its decision using evidence provided by clinicians (Bacon, 1997; Litwack et al., 1993; Cohen et al., 1978; and Palermo et al., 1991). It is precisely this approach that has been recommended in the Fallon report (1999) in the form of a Reviewable Sentence Board which he suggests might be chaired by a circuit judge and have a structure similar to a parole board.

4.9 The Limitations of this study and suggestions for future research

The conclusions of the current study are limited as a result of the number of subgroups: two types of PCL-Rs and three raters, which resulted in small sample sizes within each. The matter was further complicated by the nature of the subject records in the interview and records based PCL-Rs differing in quality. As a result of these two considerations it was not possible to ascertain if there was an effect of rater experience on the validity of the CATS or whether rater experience interacted with record quality. In addition, direct comparisons could not be made between raters as they had not completed assessments on the same subjects. Similarly, it was not possible to deduce whether the reduced validity of the CATS in the records based PCL-R sample was the result of record quality or whether the trend toward reduced validity in higher score ranges was reliable.

The trends evident in the current study suggesting an effect of rater and record quality on the validity of the CATS requires to be examined further on larger numbers of subjects, on consistently high quality records with a true comparison of raters in a within subject design. The reduced validity of the CATS for records based PCL-Rs requires further examination and should the trend be both replicated and consistent, consideration of re-weighting may be beneficial. The potential of a two pass actuarial equation to further increase accuracy should be explored.

4.10 Study Conclusions

- The VRAG calculated using the CATS in place of the PCL-R was correlated at 0.95 in a sample of 100 UK forensic inpatients.

- The VRAG loses 10% of its' predictive accuracy when calculated using the CATS, achieving 66% accuracy rather than 76% when calculated using the PCL-R a discrepancy which has operational significance.

- VRAGs calculated using the CATS do not differ from those calculated using the PCL-R by more than one scaled VRAG score.

- The CATS VRAG loses some comparative accuracy with the PCL-R VRAG when calculating the VRAG in higher score ranges.

- The CATS is a cohesive scale that primarily reflects the second factor of the PCL-R indicative of 'chronically unstable, antisocial, and socially deviant lifestyle'.

- The relevance of actuarial assessment was questioned in light of the non-uptake of those methods within all UK special hospitals despite the assertions of psychologists such as Quinsey et al. (1998):-

> "We are calling on clinicians to do risk appraisal in a new way - a way that is different from that in which most of us were trained. We are not advising the addition of actuarial methods to existing practice, but rather the complete replacement of existing practice with actuarial methods."

5 References.

American Psychiatric Association (1974). *Report of the American Psychiatric Association Task Force on clinical aspects of the violent individual 20.* On file with the American Psychiatric Association. Cited in Otto RK (1994). On the ability of mental health professionals to 'predict dangerousness': A commentary on interpretations of the 'dangerousness literature'. *Law and Psychology Review*, Vol.18. p.43-68.

American Psychological Association (1978). Report of the task force on the role of psychology in the criminal justice system. *American Psychologist*, Dec. p.1099-1113.

Apperson LJ; Mulvey EP & Lidz CW (1993). Short-term clinical prediction of assaultive behaviour: Artifacts of research methods. *American Journal of Psychiatry*, Vol.150. p.1374-1379.

Armor DJ (1974). Chapter 2: Theta reliability and factor scaling. In Costner HL (Ed). *Sociological Methodology,* 1973-1974. Jossey Bass.

Bacon P (1997). Assessing risk: are we being overcautious? *The British Journal of Psychiatry*, Vol. 170, supp 32. p.30-31.

Bailey J & MacCulloch MJ (1994). Judgements of dangerousness and release decisions in Special Hospitals. *Issues in Criminological and Legal Psychology*, Vol.21. p.76-83.

Barefoot v. Estelle (1983). 463 U.S. 880. Cited in Webster CD; Harris GT; Rice ME & Quinsey VL (1994). *The Violence Risk Prediction Scheme: Assessing Dangerousness in High Risk Men.* Centre of Criminology, University of Toronto, Canada.

Barrett P & Kline P (1982). An item and radial parcel factor analysis of the 16PF Questionnaire. *Personal and Individual Differences*, Vol.3. p.259-270.

Barrett P (Personal communication) Psychology Department, State Hospital.

Barrett PT (1996). *The PsWin Psychometrics Software Package.* Published at http://www.liv.ac.uk/ ~pbarrett/programs.htm

Belanger N & Earls C (in press) An actuarial model for the prediction of recidivism among sex offenders. *Journal of Interpersonal Violence*, Cited in Quinsey VL; Harris GT; Rice ME & Cormier CA (1998) *Violent Offenders: Appraising and Managing Risk.* American Psychological Association. Washington.

Bingley W (1997). Assessing dangerousness: protecting the interests of patients. *British Journal of Psychiatry*, Vol.170 (suppl. 32) p.28-29.

Bjorkly S (1994). The Scale for the Prediction of Aggression and Dangerousness in Psychotic Patients (PAD): A prospective pilot study. *Criminal Justice and Behaviour,* Vol.21. p.341-356.

Bjorkly S (1995). Prediction of aggression in psychiatric patients: A review of prospective prediction studies. *Clinical Psychology Review*, Vol.15. 475-502.

Black T & Spinks S (1985). Predicting outcomes of mentally disordered and dangerous offenders. In DP Farrington & R Tarling (Eds) *Prediction in criminology*. (p.176-192). Albany: State University of New York Press.

Blomhoff S; Seim S & Friis S. (1990). Can prediction of violence among psychiatric inpatients be improved? *Hospital and Community Psychiatry*, Vol.41. p. 771-775.

Bonita J; Hanson K & Law M. (1998). The prediction of criminal and violent recidivism among mentally disordered offenders: A Meta-Analysis. *Psychological Bulletin*, Vol. 123. P.123-142.

Borum R; Swartz M & Swanson J (1996). Assessing and managing violence risk in clinical practice. *Journal of Prac. Psych. And Behavioural Health.* Vol. 4. P.205-215.

Campbell JC (1995). *Assessing Dangerousness: Violence by Sexual Offenders, Batterers, and Child Abusers*. Sage Publications.

Cattell RB (1966). The Scree test for the number of factors. *Multivariate Behavioural Research*, April. P.245-276.

Chiswick D & Cope R. (Eds) (1995). *Seminars in Practical Forensic Psychiatry*. Royal College of Physicians. Gaskell.

Clanon TL & Jew C (1985). Predictions from assessments of violent offenders under stress: A Fifteen-year experience. *Criminal Justice and Behaviour*, Vol.12. p.485-499.

Cocozza J & Steadman H (1976). The failure of psychiatric predictions of dangerousness: clear and convincing evidence. *Rutgers Law Review,* Vol.29. p.1084-1101.

Cocozza JJ & Steadman HJ (1974). Some refinements in the measurement and prediction of dangerous behaviour. *American Journal of Psychiatry*, Vol.131. p.1012-1014.

Cocozza JJ; Melick ME & Steadman HJ (1978). Trends in violent crime among ex-mental patients. *Criminology,* Vol.16. p.317-334.

Cohen DA (1996). *Notes on the clinical assessment of dangerousness in offender populations.* Psychiatry on line.

Cohen ML; Groth AN & Siegal R (1978). The clinical prediction of dangerousness. *Crime and Delinquency,* Vol.24. p.28-39.

Convit A; Jaeger J; Pin Lin C; Meisner M & Volavka J (1988). Predicting assaultiveness in psychiatric inpatients: A Pilot Study. *Hospital and Community Psychiatry*, Vol.39. p.429-434.

Cooper RP & Werner PD (1990). Predicting violence in newly admitted inmates: A lens model analysis of staff decision making. *Criminal Justice and Behaviour*, Vol.17. p.431-447.

Copas JB & Whiteley JS (1976). Predicting success in the treatment of psychopaths. *British Journal of Psychiatry,* Vol.129. p.388-392.

Cunningham MD & Reidy TJ (1998). Integrating base rate data in violence risk assessments at capital sentencing. *Behavioural Sciences and the Law*, Vol.16. p.71-95.

Dawes RM; Faust D & Meehl PE (1989). Clinical versus actuarial judgement. *Science*, Vol.243. p. 1668-1674.

Diamond BL (1974). The psychiatric prediction of dangerousness. *University of Pennsylvania Law Review*, Vol.123. p.439-452.

Douglas KS; Ogloff JRP; Nicholls TL & Grant I (1998). *Assessing risk for violence among psychiatric patients: Comparison of the HCR-20 risk assessment scheme and the Psychopathy Checklist: Screening Version.* Manuscript under review. Cited in Douglas KS (1999). *HCR-20 Violence Risk Assessment Scheme: Overview and Annotated Bibliography.* http://www.sfu.ca/psychology/groups/mhlpi/hcr-20.htm

Duggan C (1997). Introduction: Special Issue, Assessing risk in the mentally disordered. *The British Journal of Psychiatry*, Vol.170, suppl. 32. p.1-3.

Ennis B & Emery R (1978a). *The Rights of Mental Patients*. New York. Avon. Cited in Monahan J (1984). The prediction of violent behaviour: Toward a second generation of theory and policy. *American Journal of Psychiatry*, Vol.141. p.10-15.

Ennis B & Emery R (1978b). *The Rights of Mental Patients*. New York. Avon. Cited in Otto RK (1992). Prediction of dangerous behaviour: A review and analysis of "second-generation" research. *Forensic Reports,* Vol.5. p. 103-133.

Fallon P; Bluglass R; Edwards B & Daniels G. (1999). *Executive summary of the report of the Committee of Inquiry into the Personality Disorder Unit, Ashworth Special Hospital.* The Stationary Office. London.

Farrington DP (1989). Early predictors of adolescent aggression and adult violence. *Violence and Victims,* Vol.4. p. 79-100.

Faust D & Ziskin J (1988). The expert witness in psychology and psychiatry. *Science*, Vol.241. p.31-35. Cited in Webster CD; Harris GT; Rice ME & Quinsey VL (1994). *The Violence Risk Prediction Scheme: Assessing Dangerousness in High Risk Men*. Centre of Criminology, University of Toronto, Canada.

Foot P (1990). *Ethics and the death penalty: Participation of forensic psychiatrists in capital trials.* In R. Rosner & R. Weinstock (Eds). Ethical practice in psychiatry and the law (pp.207-217). New York. Plenum. Cited in Grisso T & Appelbaum P (1992). Is it unethical to offer predictions of future violence? *Law and Human Behaviour*, Vol.16. p.621-633.

Forth AE; Hart SD & Hare RD (1990). Assessment of psychopathy in male young offenders. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, Vol.2. p.342-344.

Gardner W; Lidz CW; Mulvey EP & Shaw EC (1996). Clinical versus actuarial predictions of violence in patients with mental illness. *Journal of Consulting and Clinical Psychology*, Vol.64. p.602-609.

Genfreau P; Little T & Goggin C (1996). A Meta-analysis of the predictors of adult offender recidivism: What works! *Criminology,* Vol.34. p.575-607.

Glass GV & Hopkins KD (1996). *Statistical methods in education and psychology*. Allyn & Bacon.

Goldberg LR (1970). Man versus model of man: A rationale, plus some evidence, for a method of improving on clinical inferences. *Psychological Bulletin*, Vol.73. p.422-432.

Gottfredson SD & Gottfredson DM (1988). Violence prediction methods: Statistical and clinical strategies. *Violence and Victims*, Vol.3. p.303-333. Cited in Litwack TR; Kirschner SM & Wack RC. (1993). The Assessment of dangerousness and predictions of violence: Recent research and future prospects. *Psychiatric Quarterly*, Vol.64. p.245-273.

Governor's Task Force on Community Protection. Final report to Booth Gardner, Governor, State of Washington. Olympia, WA. Department of Social and Health Services. Office of the Secretary. Cited in Webster CD; Harris GT; Rice ME & Quinsey VL (1994). *The Violence Risk Prediction Scheme: Assessing Dangerousness in High Risk Men.* Centre of Criminology, University of Toronto, Canada.

Grann M; Belfrage H & Tengstr MA (1998). *Actuarial assessment of risk for violence: Predictive validity of the VRAG and the historical part of the HCR-20.* Manuscript under review. Cited in Douglas KS (1999) *HCR-20 Violence Risk Assessment Scheme: Overview and Annotated Bibliography.* http://www.sfu.ca/psychology/groups/mhlpi/hcr-20.htm

Greenland C (1980). Psychiatry and the prediction of dangerousness. *Journal of Psychiatric Treatment and Evaluation*, Vol.2. p.97-103.

Grisso T & Appelbaum P (1992). Is it unethical to offer predictions of future violence? *Law and Human Behaviour*, Vol.16. p.621-633.

Grove WM & Meehl PE (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy and Law*, Vol.2. p.293-323.

Hammond S (Personal communication) Psychology Department Broadmoor Hospital.

Hare RD (1991). *The Hare Psychopathy Checklist - Revised.* Multi-Health Systems, Inc. Toronto.

Harris G; Rice ME & Quinsey VL (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behaviour,* Vol.15. p.315-355.

Harris GT; Rice ME & Cormier CA (1991). Psychopathy and violent recidivism. *Law and Human Behaviour*, Vol.15. p.625-637.

Harris M (1997). Training trainers in risk assessment. *British Journal of Psychiatry*, Vol.170 (suppl. 32) p.35-36.

Hart SD; Hare RD & Forth AE (1994). Psychopathy as a risk marker for violence: Development and validation of a screening version of the revised Psychopathy Checklist. In Monahan J & Steadman HJ (Eds.) *Violence and Mental Disorder: Developments in Risk Assessment.* The University of Chicago Press.

Hart SD; Kropp PR & Hare RD (1988) Performance of male psychopaths following conditional release from prison. *Journal of Consulting and Clinical Psychology*. Vol.56. p.227-232.

Hedlund JL; Sletten IW; Altman SH & Evenson RC (1973). Prediction of patients who are dangerous to others. *Journal of Consulting and Clinical Psychology*, Vol.29. p.443-447.

Hodge J (Personal Communication) Psychology Department Rampton Hospital.

Hoffman PB & Beck JL (1974). Parole decision-making: A salient factor score. *Journal of Criminal Justice*, Vol.2. p.195-206.

Hoffman PB & Beck JL (1985). Recidivism among released federal prisoners: Salient factor score and five-year follow-up. *Criminal Justice and Behaviour*, Vol.12. p.501-507.

Home Office (1993). *Statistics of Mentally Disordered Offenders, England and Wales, 1991.* London: Government Statistics Service. Cited in Chiswick D & Cope R. (Eds) (1995). *Seminars in Practical Forensic Psychiatry*. Royal College of Physicians. Gaskell.

Home Office and Department of Health and Social Security (1975) *Report of the Committee on Mentally Abnormal Offenders (Butler report).* Cmnd 6244. HMSO. London. Cited in Chiswick D & Cope R. (Eds) (1995). *Seminars in Practical Forensic Psychiatry*. Royal College of Physicians. Gaskell.

Howard R & Lumsden J (1997). CNV predicts violent outcomes in patients released from special hospital. *Criminal Behaviour and Mental Health*, Vol.7. p.237-240.

Janofsky JS; Spears S & Neubauer DN (1988). Psychiatrists' accuracy in predicting violent behaviour on an inpatient unit. *Hospital and Community Psychiatry,* Vol.39. p.1090-1094.

Jennrich RI & Sampson PF (1996). Rotation for simple loadings. *Psychometricka,* Vol.31. p.313-323.

Jones T; Beidleman WB & Fowler RD (1981). Differentiating violent and nonviolent prison inmates by use of selected MMPI Scales. *Journal of Clinical Psychology*, Vol.37. p.673-678.

Kandel E; Brennan PA & Mednick SA (1989). Minor physical anomalies and recidivistic adult violent criminal behaviour. *Acta Psychiatr Scand.*, Vol.79. p.103-7.

Kirk A (1989). The prediction of violent behaviour during short-term civil commitment. *Bulletin of the American Academy of Psychiatry and Law* Vol.17. p.345-352.

Klassen D & O'Connor WA (1988a). A prospective study of predictors of violence in adult male mental health admissions. *Law and Human Behaviour*, Vol.12. p.143-158.

Klassen D & O'Connor WA (1988b). Predicting Violence in schizophrenic and non-schizophrenic patients: A prospective study. *Journal of Community Psychology*, Vol.16. p. 217-227.

Klassen D & O'Connor WA (1989). Assessing the risk of violence in released mental patients: A cross-validation study. *Psychological Assessment: A Journal of Consulting and Clinical Psychology,* Vol.1. p.75-81.

Kleinmuntz DN (1991). Decision making for professional decision makers. *Psychological Science*, Vol.2. p.135-141. Cited in Ziskin J. (1995) *Coping with Psychiatric and Psychological Testimony. Volume 1. Basic Information.* 5th edition. Law and Psychology Press. LA.

Kozol H; Boucher R & Garofalo R (1972). The diagnosis and treatment of dangerousness. *Crime and Delinquency*, Vol.18. p.371-392.

Krakowski MI & Czobor P (1994). Clinical symptoms, neurological impairment, and prediction of violence in psychiatric inpatients. *Hospital and Community Psychiatry*, Vol.45. p. 700-705.

Kroner DG & Mills JF (1997). *The VRAG: Predicting institutional misconduct in violent offenders.* Paper presented at the Annual Convention of the Ontario Psychological Association, Toronto, Canada. Cited in Quinsey VL; Harris GT; Rice ME & Cormier CA (1998). *Violent Offenders: Appraising and Managing risk.* American Psychological Association. Washington.

Lansing AE; Lyons JS; Martens LC; O'Mahoney MT; Miller SI & Obolsky A (1997). The treatment of dangerous patients in managed care. *General Hospital Psychiatry*, Vol.19. p.112-118.

Levinson RM & Ramsay G (1979). Dangerousness, stress, and mental health evaluation. *Journal of Health and Social Behaviour*, Vol.20. p.178-187.

Lidz CW; Mulvey EP & Gardner W (1993). The accuracy of predictions of violence to others. *Journal of the American Medical Association*, Vol.269. p.1007-51011.

Limandri BJ & Sheridan DJ (1995). Prediction of intentional interpersonal violence: An introduction. In Campbell JC (Ed.) *Assessing Dangerousness: Violence by Sexual Offenders, Batters, and Child Abusers*. Sage Publications, London.

Lindquist P & Allebeck P (1990). Schizophrenia and crime. A longitudinal follow up of 644 schizophrenics in Stockholm. *British Journal of Psychiatry.* Vol.157. p.345-350. Cited in Chiswick D & Cope R. (Eds) (1995). *Seminars in Practical Forensic Psychiatry.* Royal College of Physicians. Gaskell.

Litwack TR; Kirschner SM & Wack RC. (1993). The Assessment of dangerousness and predictions of violence: Recent research and future prospects. *Psychiatric Quarterly,* Vol.64. p.245-273.

Logan C (Personal communication) Psychology Department Ashworth Hospital.

MacLean (1999). *The MacLean Consultation Paper.* The Scottish Office. http:/www.scotland.gov.uk/ library/ documents-w10/mcr-00.htm

Marra HA; Konzelman GE & Giles PG (1987). A clinical strategy to the assessment of dangerousness. *International Journal of Offender Therapy and Comparative Criminology*, Vol.31. p.291-299.

McNeil DE & Binder RL (1987). Predictive validity of judgements of dangerousness in emergency civil commitment. *American Journal of Psychiatry*, Vol.144. p.197-200.

McNeil DE & Binder RL (1991). Clinical assessment of the risk of violence among psychiatric inpatients. *American Journal of Psychiatry*, Vol.148. p.1317-1321.

McNeil DE & Binder RL (1994). Screening for risk of inpatient violence: Validation of an actuarial tool. *Law and Human Behaviour*, Vol.18. p.579-586.

McNeil DE, Sandbery DA & Binder RS (1998). The relationship between confidence and accuracy in clinical assessment of psychiatric patients' potential for violence. *Law and Human Behaviour,* Vol.22. p.655-669.

McNeil DE; Binder RL & Greenfield TK (1988). Predictors of violence in civilly committed acute psychiatric patients. *American Journal of Psychiatry*, Vol.145. p.965-970.

Meehl P (1973). *Psychodiagnosis: Selected Papers*. University of Minnesota Press, Minneapolis, MN. Cited in Palermo GB; Liska FJ; Palermo MT & Forno GD (1991). On the predictability of violent behaviour: considerations and guidelines. *Journal of Forensic Sciences*, Vol.36. 5, p. 1435-1444.

Meehl PC & Rosen A (1955). Antecedent probability and the efficiency of certain psychometric signs, patterns, and cutting scores. *Psychological Bulletin*, Vol.52. p.194-216. Cited in Megargee EI (1995). Assessing and understanding aggressive and violent patients in problems in personality assessment In JN Butcher (Ed) *Personality Assessment: Practical Approaches*. Oxford University Press. p. 395-409.

Meehl PE (1954). *Clinical versus statistical prediction*. Minneapolis: University of Minnesota Press. Cited in Rice ME (1997). Violent offender research and implications for the criminal justice system. *American Psychologist,* Vol.52. 414-423.

Meehl PE (1986). Causes and effects of my disturbing little book. *Journal of Personality Assessment*, Vol.50. p.370-375.

Meehl PE (1996). *Credentialed persons, credentialed knowledge*. Paper presented at the 104th Annual Convention of the American Psychological Association. Toronto, Ontario, Canada. Cited in Rice ME (1997). Violent offender research and implications for the criminal justice system. *American Psychologist*, Vol.52. 414-423.

Meehl PE (1997). Credentialed persons, credentialed knowledge. *Clinical Psychology: Science and Practice,* Vol.4. p.91-98.

Megargee EI (1976). The prediction of dangerous behaviour. *Criminal Justice and Behaviour*, Vol.3. p.3-21.

Megargee EI (1995). Assessing and understanding aggressive and violent patients. In Butcher JN (Ed*). Clinical Personality Assessment: Practical Approaches*. Oxford University Press. p.395-409.

Melton G; Petrila J; Poythress, N & Slobogin, C. (1987). *Psychological evaluations for the courts*. New York: Guilford. Cited in Grisso T & Appelbaum P (1992). Is it unethical to offer predictions of future violence? *Law and Human Behaviour*, Vol.16. p.621-633.

Menzies R; Webster CD; McMain S; Staley S & Scaglione R (1994). The dimensions of dangerousness revisited: Assessing forensic predictions about violence. *Law and Human Behaviour*, Vol.18. p.1-28.

Menzies RJ; Webster CD & Sepejak DS (1985). Dimensions of dangerousness; Evaluating the accuracy of psychometric predictions of violence among forensic patients. *Law and Human Behaviour*, Vol.9. p.49-70.

Miller M & Morris N (1986). Predictions of dangerousness: ethical concerns and proposed limits. *Notre Dame Journal of Law and Public Policy*, Vol.2, p.393-444. Cited in Miller M & Morris N (1988). Predictions of dangerousness: An argument for limited use. *Violence and Victims*. Vol.3. p.263-283.

Miller M & Morris N (1988). Predictions of dangerousness: An argument for limited use. *Violence and Victims*. Vol.3. p.263-283.

Monahan J & Steadman H (1983). Crime and mental illness: an epidemiological approach. In *Crime and Justice*, Vol.4 (eds Morris N & Tonry M). Chicago, IL: University of Chicago Press. Cited in Wessely S (1997). The epidemiology of crime, violence and schizophrenia. *The British Journal of Psychiatry*, Vol. 170, supp 32. p.8-11.

Monahan J & Steadman H (1983). Mentally disordered offenders: Perspectives from law and social science. New York: Plenum. Cited in Borum R; Swartz M & Swanson J (1996). Assessing and Managing Violence Risk in Clinical Practice. *Journal of Prac. Psych. and Behavioural Health,* Vol.4. p.205-215.

Monahan J (1981). *The Clinical Prediction of Violent Behaviour*. Washington, DC. GPO.

Monahan J (1992). Mental disorder and violent behaviour: Perceptions and evidence. *American Psychologist*, Vol. 47. p.511-21. Cited in Borum R; Swartz M & Swanson J (1996). Assessing and managing violence risk in clinical practice. *Journal of Prac. Psych. and Behavioural Health.* Vol.4. p.205-215.

Monahan J (1996). Violence prediction: The past twenty and the next twenty years. *Criminal Justice and Behaviour*, Vol.23. p.107-120.

Moore M; Estrich S; McGillis D & Spelman W. (1983). *Dealing with dangerous offenders*. Cambridge, MA. Harvard University Press. Cited in Miller M & Morris N (1988). Predictions of dangerousness: An argument for limited use. *Violence and Victims,* Vol.3. p.263-283.

Mossman D. (1994). Assessing Predictions of Violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology,* Vol.62. 783-792.

Mullen JM & Reineher RC (1983). Predicting dangerousness of maximum security forensic mental patients. *The Journal of Psychiatry and Law,* Summer, p.223-231.

Mulvey EP & Lidz CW (1995). Conditional Prediction: A model for research on dangerousness to others in a new era. *International Journal of Law and Psychiatry*, Vol.18. p.129-143.

NHS Executive (1994). *Guidance on the Discharge of Mentally Disordered People from Hospital and their Continuing Care in the Community*. HSG (94)27. London, DoH. Cited in Chiswick D & Cope R. (Eds) (1995). *Seminars in Practical Forensic Psychiatry*. Royal College of Physicians. Gaskell.

Noak J. (1997). Assessment of the risks posed by people with mental illness. *Nursing Times*, Vol.93. p.34-36.

North East and South East Thames Regional Health Authorities (1994). *The Report of the Inquiry into the Care and Treatment of Christopher Clunis*. Chairman Jean Ritchie QC. London, HMSO. Cited in Miers D (1996). Liability for injuries caused by violent patients. *Medicine, Science and Law*, Vol.36. p.15-24.

Otto RK (1992). Prediction of dangerous behaviour: A review and analysis of "second-generation" research. *Forensic Reports,* Vol.5. p. 103-133.

Otto RK (1994). On the ability of mental health professionals to 'predict dangerousness': A commentary on interpretations of the 'dangerousness literature'. *Law and Psychology Review*, Vol.18. p.43-68.

Palermo GB; Liska FJ; Palermo MT & Forno GD (1991). On the predictability of violent behaviour: considerations and guidelines. *Journal of Forensic Sciences*, Vol.36. p. 1435-1444.

Pollock N & Webster C (1990). The clinical assessment of dangerousness. In Bluglass R & Bowden P (Eds) *Principles and Practice of Forensic Psychiatry*. Churchill Livingstone. London.

Pollock NL (1990). Accounting for predictions of dangerousness. *International Journal of Law and Psychiatry*, Vol.13. p.207-215.

Polythress NG (1992). Expert testimony on violence and dangerousness: Roles for mental health Professionals. *Forensic Reports*, Vol.5. p.135-150. Cited in Cunningham MD & Reidy TJ (1998) Integrating base rate data in violence risk assessments at capital sentencing. *Behavioural Sciences and the Law,* Vol.16. p.71-95.

Prins H. (1996a). Risk assessment and management in the criminal justice and psychiatry. *Journal of Forensic Psychiatry*, Vol.7. p.42-62. Cited in Noak J. (1997). Assessment of the risks posed by people with mental illness. *Nursing Times*, Vol.93. p.34-36.

Prins H. (1996b). Risk assessment and management in the criminal justice and psychiatry. *Journal of Forensic Psychiatry*, Vol.7. p.42-62. Cited in Bluglass R & Bowden P (Eds) *Principles and Practice of Forensic Psychiatry*. Churchill Livingstone. London.

Quinsey VL & Ambtman R (1979). Variables affecting psychiatrists' and teachers' assessment of the dangerousness of mentally ill offenders. *Journal of Consulting and Clinical Psychology,* Vol.47. p.353-362.

Quinsey VL & MaGuire A (1986). Maximum security psychiatric patients: Actuarial and clinical prediction of dangerousness. *Journal of Interpersonal Violence*, Vol.1. p.143-171.

Quinsey VL (1979). Assessment of the dangerousness of mental patients held in maximum security. *International Journal of Law and Psychiatry*, Vol.2. p.389-406.

Quinsey VL; Harris GT; Rice ME & Cormier CA (1998). *Violent Offenders: Appraising and Managing Risk.* American Psychological Association. Washington.

Quinsey VL; Rice ME & Harris GT (1995). Actuarial prediction of sexual recidivism. *Journal of Interpersonal Violence*, Vol.10. 85-105.

Quinsey VL; Skilling T & Rougier-Chapman C (1997). *Inter-Ministerial Population Review Initiative: Final Report.* Toronto, Ontario, Canada: Ministry of Community and Social Services. Cited in Quinsey VL; Harris GT; Rice ME & Cormier CA (1998). *Violent Offenders: Appraising and Managing Risk.* American Psychological Association. Washington.

Reed J (1992). *Review of health and social services for mentally disordered offenders and others requiring similar services. Final Summary Report.* HMSO. London.

Reed J (1994). *Report of the Department of Health and Home Office working group on psychopathic disorder.* Department of Health, HMSO.

Reed J. (1997). Risk assessment and clinical risk management: the lessons from recent enquiries *The British Journal of Psychiatry*, Vol. 170, suppl. 32. p.4-7.

Rice ME & Harris GT (1995b). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology*, Vol.63. p.737-748.

Rice ME & Harris GT (1997). Cross validation and extension of the Violence Risk Appraisal Guide with child molesters and rapists. *Law and Human Behaviour,* Vol.21. p.231-241.

Rice ME (1997). Violent offender research and implications for the criminal justice system. *American Psychologist*, Vol.52. 414-423.

Rice ME (1999). Personal communication, email, with Barrett P.

Rofman ES; Askinazi C & Fant E. (1980). The prediction of dangerous behaviour in emergency civil commitment. *American Journal of Psychiatry*, Vol.137. p. 1061-1064.

Ross DJ; Hart SD & Webster CD (1998). *Facts and fates: Testing the HCR-20 against aggressive* behaviour in hospital and community. Unpublished manuscript. Cited in Douglas KS (1999). *HCR-20 Violence Risk Assessment Scheme: Overview and Annotated Bibliography* http://www.sfu.ca/psychology/groups/mhlpi/hcr-20.htm

Royal College of Psychiatrists (1996). *Report of the confidential enquiry into homicides and suicides by mentally ill people.* Royal College of Psychiatrists, London.

Russo G (1994). Follow-up of 91 mentally ill criminals discharged from the maximum security hospital in barcelona P.G. *International Journal of Law and Psychiatry,* Vol. 17. p.279-301.

Sackett DL; Haynes RB; Guyatt GH & Tugwell P (1985). *Clinical Epidemiology: A Basic Science for Clinical Medicine.* Little, Brown & Company, Boston.

Scott PD (1977). Assessing dangerousness in criminals. *British Journal of Psychiatry*, Vol.131. p.127-142.

Scottish Office (1999). *Health, social work and related services for mentally disordered offenders in Scotland.* Department of Health, HMSO.

Selby MJ (1984). Assessment of violence potential using measures of anger, hostility and social desirability. *Journal of Personality Assessment*, Vol. 48. P. 531-544.

Sepejak D; Menzies RJ; Webster CD & Jensen FAS (1983). Clinical predictions of dangerousness: Two-year follow-up of 408 pre-trial forensic cases. *Bulletin of the American Academy of Psychiatry and Law*, Vol.11. p. 171-181.

Serin RC & Amos NL (1995). The role of psychopathy in the assessment of dangerousness. *International Journal of Law and Psychiatry,* Vol.18. p.231-238.

Shah, SA (1978). Dangerousness: A paradigm for exploring some issues in law and psychology. *American Psychologist*, March, p.224-238.

Smith SM (1993). The prediction of dangerous behaviour. In P.J. Resnick (Ed.) *Forensic Psychiatry Review Course,* American Academy of Psychiatry and the Law, p.539-541. Cited in Cunningham MD & Reidy TJ (1998). Integrating base rate data in violence risk assessments at capital sentencing. *Behavioural Sciences and the Law*, Vol.16. p.71-95.

Snowden P (1997). Practical aspects of clinical risk assessment and management. *The British Journal of Psychiatry,* Vol. 170, supp 32. p. 32-34.

Steadman H & Cocozza J (1974). *Careers of the Criminally Insane.* Lexington, Mass. Lexington Books.

Steadman H (1977). A new look at recidivism among Patuxent inmates. *Bulletin of the American Academy of Psychiatry and Law,* Vol.5. p.200-209.

Steadman HJ & Cocozza J (1978). Criminology: Psychiatry, dangerousness and the repetitively violent offender. *The Journal of Criminal Law and Criminology,* Vol.69. p.226-231.

Steadman HJ & Morrissey JP (1982). Predicting violent behaviour: A note on a cross-validation study. *Social Forces*, Vol.61. p.475-483.

Stone A (1985). The new legal standard of dangerousness: Fair in theory, unfair in practice. In C.D. Webster, MH Ben-Aron & SJ Hucker (Eds). *Dangerousness: Probability and prediction, psychiatry and public policy* (p.13-24). New York, NY: Cambridge University Press. Cited in Pollock N & Webster C (1990). The clinical assessment of dangerousness. In Bluglass R & Bowden P (Eds) *Principles and Practice of Forensic Psychiatry*. Churchill Livingstone. London.

Strand S; Belfrage, H; Fransson G & Levander S (1998). *Clinical and risk management factors in risk prediction of mentally disordered offenders: More important than actuarial data? Manuscript under reivew.* Cited in Douglas KS (1999). *HCR-20 Violence Risk Assessment Scheme: Overview and Annotated Bibliography*. Http://www.sfu.ca/psychology/groups/mhlpi/hcr-20.htm

Swanson JW (1994). Mental disorder, substance abuse, and community violence: A epidemiological approach. In Monahan J & Steadman H (Eds) *Violence and Mental Disorder: Developments in Risk Assessment.* Chicago. University of Chicago Press. p.101-136.

Swanson JW; Holzer CE; Ganju VK & Jono RT (1990). Violence and psychiatric disorder in the community: Evidence from the Epidemiologic Catchment Area Surveys. *Hospital and Community Psychiatry,* Vol.41. p.761-770. Cited in Borum R; Swartz M & Swanson J (1996). Assessing and managing violence risk in clinical practice. *Journal of Prac. Psych. and Behavioural Health,* Vol.4. p.205-215.

Swett C & Hartz SC (1984). Antecedents of violent acts in a prison hospital. *The American Journal of Social Psychiatry,* Vol.4. p. 24-29.

The State Hospitals Board for Scotland (1999). *Health Improvement Programme.*

Thornberry TP & Jacoby JE (1979). *The Criminally Insane: A Community Follow-up of Mentally Ill Offenders.* Chicago. University of Chicago Press.

Turk DC & Salovey P (1988). Cognitive Structures, cognitive processes, and cognitive-behaviour modification: II. Judgements and inferences of the clinician. *Cognitive Therapy and Research,* Vol.9. p.19-33. Cited in Mulvey EP & Lidz CW (1995) Conditional Prediction: A model for research on dangerousness to others in a new era. *International Journal of Law and Psychiatry*, Vol.18. p.129-143.

Velicer WF (1976). Determining the number of components from the matrix of partial correlations. *Psychometrik,*. Vol.41. p.321-327.

Walker N. (1978). Dangerous people. *International Journal of Law and Psychiatry*. Vol.11. p.37-50. Cited in Chiswick D & Cope R. (Eds) (1995). *Seminars in Practical Forensic Psychiatry*, Royal College of Physicians. Gaskell.

Webster CD; Harris GT; Rice ME & Quinsey VL (1994). *The Violence Risk Prediction Scheme: Assessing Dangerousness in High Risk Men.* Centre of Criminology, University of Toronto, Canada.

Webster CD; Sepejak DS; Menzies RJ; Slomen DJ; Jensen FAS & Butler BT (1984). The reliability and validity of dangerous behaviour predictions. *Bulletin of the American Acad*emy of *Psychiatry and Law,* Vol.12. p.41-50.

Webster D; Douglas KS; Eaves SD & Hart SD (1997). Assessing risk of violence to others. In C.D. Webster & MA Jackson (Eds), *Impulsivity: Theory, assessment, and treatment*, p251-277. New York: Guilford. Cited in Douglas KS (1999). *HCR-20 Violence Risk Assessment Scheme:* Overview and Annotated Bibliography. http://www.sfu.ca/psychology/groups/mhlpi/hcr-20.htm

Werner PD & Meloy JR (1992). Decision making about dangerousness in releasing patients from long-term psychiatric hospitalization. *The Journal of Psychiatry and Law,* Spring. P.35-47.

Werner PD; Rose TL & Yesavage JA (1983). Reliability, accuracy, and decision-making strategy in clinical predictions of imminent dangerousness. *Journal of Consulting and Clinical Psychology,* Vol.51. p.815-825.

Wessely S (1997). The epidemiology of crime, violence and schizophrenia. *The British Journal of Psychiatry,* Vol.170, suppl. 32. P.8-11.

Wessely S; Castle D; Douglas AJ et al. (1994). The criminal careers of incident cases of schizophrenia. *Psychological Medicine*, Vol.24. p.483-502. Cited in Chiswick D & Cope R. (Eds) (1995). *Seminars in Practical Forensic Psychiatry*. Royal College of Physicians. Gaskell.

Yesavage JA; Werner PD; Becker JMT & Mills MJ (1982). Short-Term Civil Commitment and the Violent Patient. *American Journal of Psychiatry,* Vol.139. p.1145-1149.

Zeiss RA; Tanke ED; Fenn HH & Yesavage JA (1996). Dangerousness commitments: Indices of future violence potential? *Bulletin of the American Academy of Psychiatry and Law,* Vol. 24(2), p.247-253.

Ziskin J. (1995). *Coping with Psychiatric and Psychological Testimony. Volume 1. Basic Information.* 5th edition. Law and Psychology Press. LA.

Appendix 1

The Violence Risk Appraisal Guide

VRAG Data Collection Form **Patient no.** ☐☐☐☐

☐ 1) Lived with both biological parents to age 16?
(Disregard if parent dead)

-2 = **Yes**
+3 = **No**

.....
.....

☐ 2) Elementary (Primary) school maladjustment?

-1 = **No problems**
+2 = **Slight or moderate problems** (minor discipline or attendance)
+5 = **Severe problems** (frequent disruptive behaviour and/or attendance or behaviour resulting in expulsion or serious suspension)

.....
.....

☐ 3) History of alcohol problems?

(Parental alcoholism ☐
Teenage alcohol problem ☐
Adult (>18yrs) alcohol problem ☐
Alcohol involved in a prior offence ☐
Alcohol involved in index offence ☐)
TOTAL ☐

-1 = **Total = 0**
0 = **Total 1 or 2**
+1 = **Total = 3**
+2 = **Total 4 or 5**

.....
.....

☐ 4) Marital status?

-2 = **Ever married** (or lived common law in the same home for at least 6 months)
+1 = **Never married.**

.....
.....

☐ 5) Criminal History Score for nonviolent offences.

7- (Robbery (bank, store)
3- Robbery (purse snatching)
5- Arson & fire setting (church, house, barn)
2- Arson & fire setting (garbage can)
3- Threatening with a weapon
2- Threatening (uttering threats)
5- Theft > $1,000 inc car theft & possession of stolen property)
5- Mischief to public or private property over $1,000
2- Break & enter & commit indictable offence (burglary)
1- Theft < (inc possession of stolen goods under) $1,000
1- Mischief to public or private property < $1,000 (+public mischief)
1- Procuring a person for, or living on the avails of prostitution
1- Break & enter (inc.s break & enter with intent)
5- Fraud (extortion, embezzlement)
1- Fraud (forged cheque, impersonation)
1- Possession of a prohibited or restricted weapon
1- Trafficking in narcotics
1- Dangerous driving, impaired driving (eg. drunk)
1- Obstructing peace officer (inc. resisting arrest)
1- Causing a disturbance
1- Wearing a disguise with the intent to commit an offence

TOTAL SCORE = ☐

-2 = **Total score = 0**
0 = **Total score 1 or 2**
+3 = **Total score 3 or more**

.....

.....

☐ 6) Failure on prior conditional release (includes parole or probation violation or revocation, failure to comply, bail violation, and any new arrest while on conditional release). (Does not include absconding whilst on a supervised day-trip or escaping from an instituition).

0 = **No**
+3 = **Yes**

.....

.....

☐ 7) Age at index offence (at most recent birthday). If >1 charge at once take earliest, if single current offence brings to light previous just take current).

-5 = **=>39yrs**
-2 = **34-38yrs**
-1 = **28-33yrs**
0 = **27yrs**
+2 = **=<26yrs**

.....

.....

☐ 8) Victim injury (for index offence; the most serious is scored).

-2 = **Death**
0 = **Hospitalised**
+1 = **Treated and released**
+2 = **None or slight** (includes no victim)

.....
.....

☐ 9) Any female victim (for index offence)?

-1 = **Yes**
+1 = **No** (inc. no victim)

.....
.....

☐ 10) Meets DSM III criteria for any personality disorder?

-2 = **No**
+3 = **Yes**

.....
.....

☐ 11) Meets DSM III criteria for schizophrenia?

-3 = **Yes**
+1 = **No**

.....
.....

☐ 12) Psychopathy Checklist score?

-5 = **=<4**
-3 = **5-9**
-1 = **10-14**
0 = **15-24**
+4 = **25-34**
+12 = **=>35**

.....
.....

PCL-R VRAG Total Score ☐☐

Appendix 2

The Child and Adolescent Taxon Scale

CATS Data Collection Form

☐ 1) Elementary school maladjustment score greater than 2?
See VRAG Q2.

+1 = **Yes**
0 = **No**

.....

.....

☐ 2) Teenage alcohol problem?
See VRAG Q3.

+1 = **Yes**
0 = **No**

☐ 3) Childhood aggression rating?

1 = No evidence of aggression.
4 = Occasional moderate aggression.
7 = Occasional or frequent extreme aggression.

+1 = **If score more than 4**
0 = If less than 4.

☐ 4) More than 3 DSM-III Conduct Disorder symptoms recorded?

A) A repetitive & persistent pattern of aggressive conduct in which the basic rights of others are violated, as manifested by either of the following:
   1) Physical violence against persons or property (not to defend someone else or oneself), e.g. vandalism, rape, breaking and entering, fire-setting, mugging, assault.
   2) Thefts outside the home involving confrontation with the victim (e.g. extortion, purse-snatching, armed robbery).
   3) Chronic violations of a variety of important rules (that are reasonable and age-appropriate for the child) at home or at school (e.g. persistent truancy, substance abuse).
   4) Repeated funning away from home overnight.
   5) Persistent serious lying in and out of the home.
   6) Stealing not involving confrontation with a victim.

B) Failure to establish a normal degree of affection, empathy, or bond with others as evidenced by no more than one of the following indications of social attachment:
   1) Has one or more peer-group friendships that have lasted over six months.
   2) Extends himself or herself for others even when no immediate advantage is likely.
   3) Apparently feels guilt or remorse when such a reaction is appropriate (not just when caught or in difficulty).
   4) Avoids blaming or informing on companions.
   5) Shares concern for the welfare of friends or companions.

C) Duration of pattern of aggressive conduct of at least six months.

D) If 18 or older, does not meet the criteria for Antisocial Personality Disorder.

+1 = **Yes**
0 = **No**

☐ 5) Ever suspended or expelled from school?

+1 = **Yes**
0 = **No**

☐ 6) Arrested under the age of 16 years?

+1 = **Yes**
0 = **No**

☐ 7) Parental alcoholism?
See VRAG Q3.

+1 = **Yes**
0 = **No**

☐ 8) Lived with both parents to age 16 (except for death of parents)?
See VRAG Q3.

0 = **Yes**
+1 = **No**

Assigning Predictive Weights:

CATS scores of 0 or 1	= -3
CATS score of 2 or 3	= 0
CATS score of 4	= +2
CATS scores of 5 or more	= +3

CATS Total Score ☐☐

CATS- VRAG Total Score ☐☐

Appendix 3

The Violence Risk Appraisal Guide
Scoring Guide

VRAG SCORING GUIDE

1) Lived with both biological parents to age 16?

- This item refers to biological parents.
- If absent from home for a period exceeding one month score as negative.
- Circumstances counted as living with parents:
    - Holidays
    - Boarding school
- Circumstances counted as not living with parents:
    - D List schools
    - Parent as hospital inpatient

2) Elementary School Maladjustment.

- If not commented upon, assume absent.
- Referral to D list school is not scored positively without further information.

3) History of Alcohol Problems

- If file comments on alcohol use inconclusive look for evidence of a alcohol related dysfunction
- Score item positively if words/ phrases in notes such as "Abuse, problem, lifestyle, binge drinking, regularly intoxicated, when drunk acts aggressively".
- Parental Alcoholism refers to means Biological Parents only.

4) Marital Status

5) Criminal History Score for nonviolent offences.

- Only score positively convictions exactly equivalent to those listed.
- If in doubt as to the grade of offence give the lower score.

6) Failure on prior conditional release.

- Includes parole or probation violation or revocation, failure to comply, bail violation, and any new arrest while on conditional release.
- Does not include absconding whilst on a supervised day-trip or escaping from an institution.
- Re-admission constitutes failure unless previously given an absolute discharge.
- If previous admission was whilst awaiting trial and patient was subsequently found not-guilty and released, re-admission does not constitute failure.

Identifying the Index Offence

- If they have been transferred from prison, then take the offence that took them to prison.
- If admitted for a criminal act but never charged or convicted (e.g. Insane in bar of trial) then the criminal act is the index offence.
- If transferred from Hospital for an act that would normally result in criminal charges then score as the index offence.
- If re-admitted due to deterioration in Mental Health, then score the original index offence.

7) Age at index offence (at most recent birthday).

- If a single current offence brings to light previous just take current
- If no Index Offence then count it as age they were transferred to the State Hospital.
- If they have had multiple admissions, then take their age on their original admission to the State Hospital.

8) Victim injury

- If multiple victims the most serious injury is scored.
- If severity unclear score at the lower grade.

9) Any female victim (for index offence)?

9) Meets DSM III criteria for any personality disorder?

- Score positively if medical records specify the diagnosis.

10) Meets DSM III criteria for schizophrenia?

- Score positively if medical records specify the diagnosis.

12) Psychopathy Checklist score?

The VRAG scoring guide is the product of joint meetings between Dr. P Barrett and the raters SOR, BH and KR. The above guide is based on the notes compiled from these meetings by KR.

Appendix 4

The Violence Risk Appraisal Guide
Response Frequencies

The distribution of responses to each VRAG variable for interview and records based PCL-Rs and for the sample as a whole.

Variable	Response Options	Frequency Count			Percentage		
		Type of PCL-R					
		Int	Rec	All	Int	Rec	All
Lived with both biological parents to age 16?	Yes	22	22	44	38	52	44
	No	36	20	56	62	48	56
Elementary school maladjustment?	No problems	32	23	55	55	55	55
	Slight or moderate problems	6	7	13	10	17	13
	Severe problems	20	12	32	34	29	32
History of alcohol problems?	Cumulative Total: 0	7	13	20	12	31	20
	1	16	12	28	28	29	28
Parental alcoholism	2	14	6	20	24	14	20
Teenage alcohol problem	3	10	6	16	17	14	16
Adult alcohol problem	4	9	4	13	16	10	13
Alcohol involved in prior offence	5	2	1	3	3	2	3
Alcohol involved in prior offence							
Marital Status	Ever married?	24	11	35	41	74	35
	Never married?	34	31	65	59	26	65
Criminal History Score for non-violent offences	Total score: 0	8	9	17	14	21	17
	1 or 2	4	3	7	7	7	7
	3 or more	46	30	76	79	71	76
Failure on prior conditional release?	Yes	30	18	48	52	43	48
	No	28	24	52	48	57	52
Age at index offence?	=<26 years	31	24	55	53	57	55
	27 years	4	2	6	7	5	6
	28-33 years	13	6	19	22	14	19
	34-38 years	6	3	9	10	7	9
	=>39 years	4	7	11	7	17	11
Victim injury for index offence?	Death	18	10	28	31	24	28
	Hospitalised	16	6	22	28	14	22
	Treated and released	7	3	10	12	7	10
	None or slight	17	23	40	29	55	40
Any female victim?	Yes	20	20	40	34	48	40
	No	38	22	60	66	52	60
Meets DSM III criteria for any personality disorder?	Yes	24	16	40	41	38	40
	No	34	26	60	59	62	60
Meets DSM III criteria for schizophrenia?	Yes	48	29	77	83	69	77
	No	10	13	23	17	31	23
Psychopathy Checklist Score	=<4	2	1	3	3	2	3
	5-9	6	6	12	10	14	12
	10-14	13	5	18	22	12	18
	15-24	28	20	48	48	48	48
	25-34	9	10	19	16	24	19
	=>35	0	0	0	0	0	0

Int = Interview based PCL-Rs. Rec = Records based PCL-Rs. All = Refers to the combined sample of PCL-Rs.

Appendix 5

The Child and Adolescent Taxon Scale Response Frequencies

Response frequency on the CATS for interview based and records based

Variable	Response Options	Frequency Count Type of PCL-R			Percentage Type of PCL-R		
		Int	Rec	All	Int	Rec	All
Elementary school maladjustment score greater than 2 (on VRAG)?	Yes	20	12	32	34	29	32
	No	38	30	68	66	71	68
Teenage alcohol problem?	Yes	22	9	31	38	21	31
	No	36	33	69	62	79	69
Childhood aggression rating? 1-No evidence of aggression. 4- Occasional moderate aggression. 7-Occasional or frequency extreme aggression.	Score > 4	23	7	30	60	17	30
	Score < 5	35	35	70	40	83	70
More than 3 DSM-III Conduct Disorder symptoms recorded?	Yes	25	20	45	43	48	45
	No	33	22	55	57	52	55
Ever suspended or expelled from school?	Yes	31	16	47	53	38	47
	No	27	26	53	47	62	53
Arrested under the age of 16 years?	Yes	23	16	39	40	38	39
	No	35	26	61	60	62	61
Parental alcoholism?	Yes	30	13	43	52	31	43
	No	28	29	57	48	69	57
Lived with both parents to age 16?	Yes	21	22	43	36	52	43
	No	37	20	57	64	48	57

Int = Interview based PCL-Rs. Rec = Records based PCL-Rs. All = Refers to the combined sample of PCL-Rs.

Appendix 6

The Psychopathy Checklist - Revised
Response Frequencies

Response frequency on the PCL-R for the entire sample and subdivided into interview based and records based PCL-Rs.

Variable	Response Options	Frequency Count			Percentage		
		Int	Rec	All	Int	Rec	All
Glibness / Superficial charm	No	40	29	69	70	69	69
	Maybe	17	8	25	29	19	25
	Yes	1	5	6	2	12	6
	Omit	0	0	0	0	0	0
Grandiose sense of self worth	No	33	28	61	57	67	61
	Maybe	17	8	25	29	19	25
	Yes	8	6	14	14	14	14
	Omit	0	0	0	0	0	0
Need for stimulation / Proness to boredom	No	30	17	47	52	40	47
	Maybe	17	19	36	29	45	36
	Yes	11	6	17	19	14	17
	Omit	0	0	0	0	0	0
Pathological lying	No	37	24	61	64	57	61
	Maybe	16	11	27	28	26	27
	Yes	5	7	12	9	17	12
	Omit	0	0	0	0	0	0
Conning / Manipulative	No	39	21	60	67	50	60
	Maybe	13	7	20	22	17	20
	Yes	6	14	20	10	33	20
	Omit	0	0	0	0	0	0
Lack of remorse or guilt	No	30	8	38	52	19	38
	Maybe	15	11	26	26	26	26
	Yes	13	22	35	22	52	35
	Omit	0	1	1	0	2	1
Shallow affect	No	18	10	28	31	24	28
	Maybe	31	9	40	53	21	40
	Yes	9	23	32	16	55	32
	Omit	0	0	0	0	0	0
Callous / Lack of empathy	No	29	10	39	50	24	39
	Maybe	23	12	35	40	29	35
	Yes	6	20	26	10	48	26
	Omit	0	0	0	0	0	0
Parasitic Lifestyle	No	8	4	12	14	10	12
	Maybe	17	8	25	29	19	25
	Yes	33	30	63	57	71	63
	Omit	0	0	0	0	0	0
Poor behavioural controls	No	14	1	15	24	2	15
	Maybe	25	15	40	43	36	40
	Yes	19	26	45	33	62	45
	Omit	0	0	0	0	0	0

Int = Interview based PCL-Rs. Rec = Records based PCL-Rs. All = Refers to the combined sample of PCL-Rs.

Response frequency on the PCL-R for the entire sample and subdivided into interview based and records based PCL-Rs (continued).

Variable	Response Options	Frequency Count			Percentage		
		Int	Rec	All	Int	Rec	All
Promiscuous sexual behaviour	No	19	28	47	33	67	47
	Maybe	7	5	12	12	12	12
	Yes	30	9	39	52	21	39
	Omit	2	0	2	3	0	2
Early behavioural problems	No	23	16	39	40	38	39
	Maybe	6	6	12	10	14	12
	Yes	29	18	47	50	43	47
	Omit	0	2	2	0	5	2
Lack of realistic, long term goals	No	25	3	28	43	7	28
	Maybe	27	15	42	47	36	42
	Yes	6	22	28	10	52	28
	Omit	0	2	2	0	5	2
Impulsivity	No	8	1	9	14	2	9
	Maybe	14	16	30	24	38	30
	Yes	36	25	61	62	60	61
	Omit	0	0	0	0	0	0
Irresponsibility	No	14	6	20	24	14	20
	Maybe	37	16	53	64	38	53
	Yes	7	20	27	12	48	27
	Omit	0	0	0	0	0	0
Failure to accept responsibility for own actions	No	24	13	37	41	31	37
	Maybe	30	10	40	52	24	40
	Yes	4	19	23	7	45	23
	Omit	0	0	0	0	0	0
Many short-term marital relationships Under age 30: 0 = 0-1 1 = 2 2 = 3+ Aged 30+: 0 = 0-2 1 = 3 2 = 4+	0	40	41	81	69	98	81
	1	8	1	9	14	2	9
	2	10	0	10	17	0	10
	Omit	0	0	0	21	0	0
Juvenile Delinquency	No offences	22	25	47	38	60	47
	Minor	8	2	10	14	5	10
	Major	28	15	43	48	36	43
	Omit	0	0	0	0	0	0
Revocation of conditional release	No violations	31	31	62	53	74	62
	Minor	8	4	12	14	10	12
	Major	19	7	26	33	17	26
	Omit	0	0	0	0	0	0
Criminal versatility	0=0-3	10	29	39	17	69	39
	1=4-5	20	9	29	34	21	29
	2=6 or more	28	4	32	48	10	32
	Omit	0	0	0	0	0	0

Int = Interview based PCL-Rs. Rec = Records based PCL-Rs. All = Refers to the combined sample of PCL-Rs.